



INSTITUT
de
STATISTIQUE
de l'UNESCO

ISU/OLOM/CLOM/2014

Montréal, Septembre 2014

Enquête 2014
Observatoire des Mesures des Acquis de
l'Apprentissage

**MANUEL
POUR COMPLÉTER LE CATALOGUE DES
MESURES DES ACQUIS DE
L'APPRENTISSAGE**

CONTENUS

1. INTRODUCTION	4
2. COUVERTURE DU QUESTIONNAIRE	6
3. COMMENT COMPLÉTER LE QUESTIONNAIRE	7
4. INSTRUCTIONS POUR CHAQUE QUESTION	8
SECTION 2 : LISTE DES EVALUATIONS	8
SECTION 3: CHAMP D'APPLICATION, OBJECTIFS, FINANCEMENT ET PARTIES PRENANTES.....	11
SECTION 4: CONCEPTION ET ADMINISTRATION DU TEST	13
SECTION 5 : COUVERTURE ET PLAN D'ECHANTILLONNAGE	16
SECTION 6: TRAITEMENT ET EVALUATION DES DONNEES	18
SECTION 7: MESURES ET RESULTATS	19
SECTION 8: DISSEMINATION DES DONNEES.....	22
5. GLOSSAIRE ET DÉFINITIONS	23

Consignes d'ordre technique à respecter pour remplir les tableaux

1) Consignes de codage des données manquantes

Il est essentiel pour l'intégrité des données de respecter certains signes conventionnels pour coder les données manquantes. Il y a lieu de pouvoir distinguer les raisons auxquelles imputer le manque de données dans une catégorie spécifique dans les analyses et les rapports.

Ne laissez aucune cellule vide dans les tableaux. L'un des quatre codes conventionnels suivants doit obligatoirement être indiqué dans toute cellule sans donnée valide :

i) Catégorie sans objet = **a**

Si une catégorie ne s'applique pas à une évaluation donnée ou un examen public national de votre pays, indiquez le code « **a** » dans toutes les cellules se rapportant à cette catégorie. Ce code signifie qu'il n'existe pas de données pour les catégories concernées.

ii) Valeur nulle = **n**

Si une valeur est nulle ou négligeable, indiquez le code « **n** » dans la cellule correspondante. Si une catégorie est sans objet, et donc associée à une valeur nulle, n'indiquez pas le code « **n** », mais le code « **a** », pour catégorie sans objet. N'utilisez pas le zéro (« 0 ») pour indiquer une valeur nulle.

iii) Données incluses dans une autre catégorie = **x**

S'il n'est pas possible de désagréger une donnée pour obtenir celle qui est requise dans une cellule donnée, indiquez le code « **x** » dans la cellule dont la donnée est manquante. **Veillez préciser** dans les métadonnées la catégorie dans laquelle cette donnée est incluse si c'est le cas.

iv) Données non disponibles = **m**

En l'absence de données dans certaines catégories et pour autant que ces données ne soient pas incluses dans d'autres cellules du tableau, veuillez indiquer le code « **m** » dans toutes les cellules se rapportant à ces catégories (même si ces données auraient en principe pu être recueillies).

v) Données confidentielles = **c**

Indiquez le code « **c** » pour signaler des données que le pays ne souhaite pas rendre publiques.

2) Remarque sur les estimations ou les chiffres provisoires (*)

Veillez faire précéder les estimations ou les chiffres provisoires d'un astérisque (*).

3) Remarque sur les données partielles = **p**

Veillez faire précéder les données partielles du code « **p** ».

1. INTRODUCTION

L'institut des Statistiques de l'UNESCO (ISU) a lancé une nouvelle initiative globale visant à cartographier les différentes évaluations auxquelles les pays ont recours en vue d'assurer le suivi des acquis chez les apprenants. Les consultations internationales relatives à l'éducation ont bien montré que, outre le fait qu'il assure un accès équitable à une éducation de qualité, l'apprentissage joue un rôle fondamental au regard des perspectives de développement post-2015. Dans le domaine de l'apprentissage, on considère qu'il est important que les enfants et les jeunes acquièrent certaines aptitudes et compétences afin d'être à même de mener des vies productives et de relever les défis de l'avenir. Les évaluations nationales sont courantes dans la plupart des pays, que les enjeux soient élevés ou faibles. La documentation, l'analyse et la diffusion de la masse de connaissances que ces évaluations procurent figurent parmi les principaux objectifs du catalogue des mesures des résultats de l'apprentissage de l'ISU. L'ISU diffusera le contenu du catalogue au travers de sa base de données internationale. Le [site web de l'ISU](#) offre des informations complémentaires.

Le catalogue sera un inventaire global de toutes les évaluations nationales et internationales, et examens publics nationaux à grande échelle que les pays utilisent en vue de mesurer et faire le suivi des acquis des apprentissages à l'enseignement primaire et à l'enseignement secondaire au premier cycle (niveaux 1 et 2 de la CITE 2011). Le catalogue sera produit sur la base des données soumises à partir du questionnaire qui a été confectionné en vue de collecter des informations sur les caractéristiques des évaluations des apprenants qui sont disponibles au moment de l'enquête. Ce questionnaire compile les données des évaluations récentes à couverture nationale qui ont été administrées dans le pays entre 2000 et 2014. Les données des évaluations menées avant 2000 sont exclues du présent catalogue.

Produits. L'ensemble des informations relatives aux évaluations des apprenants générées au travers du catalogue viendra enrichir la base de données internationale de l'ISU, et sera donc à accès public. Les informations du catalogue seront combinées aux données déjà disponibles dans la base de données internationale de l'ISU en vue de produire un bref résumé par pays qui dépeindra les efforts déployés au niveau national pour mesurer et contrôler les acquis scolaires chez les apprenants. Il comprendra la liste des évaluations - nationales ou internationales, avec des enjeux élevés ou faibles- les domaines qu'elles couvrent et la mesure des acquis des apprenants dans ces différents domaines. Les résumés par pays procureront des informations clés sur ces évaluations, et sur les différences et les points communs entre les caractéristiques de l'évaluation, sans comparer la qualité des évaluations à l'échelle nationale ou d'un pays à l'autre.

Résultats escomptés. Les résumés par pays peuvent être utilisés dans le cadre de discussions éclairées sur l'opportunité d'utiliser une même évaluation par plusieurs parties prenantes. Les autorités et les établissements d'enseignement locaux peuvent recourir aux résumés en tant que sources d'information contextuelle pertinente dans le but d'examiner les caractéristiques de leurs évaluations par rapport à des objectifs spécifiques. La communauté de développement, comme les organismes d'aide bilatérale et multilatérale, pourrait envisager les résumés par pays comme faisant partie d'une boîte à outils destinée généralement à évaluer la pertinence des propositions visant à soutenir les évaluations des résultats de l'apprentissage. L'ISU veillera activement à encourager l'utilisation des résumés par pays et de l'information contenue dans la base de données internationale ; par exemple, afin de documenter les révisions du champ d'application et de la couverture des évaluations auprès des apprenants qui sont initiées par les pays eux-mêmes.

Bénéfices pour les pays. Les premières estimations de l'ISU indiquent que plus de 150 pays mesurent actuellement les niveaux d'apprentissage à l'aide d'évaluations et d'examens nationaux ou participent à des programmes d'évaluation internationaux (incluent régionaux ou transnationaux). Jusqu'à présent, aucun effort global n'a été entrepris en vue de rassembler systématiquement des informations concernant les choix des pays et leurs liens avec l'agenda sur les acquis de l'apprentissage. Le catalogue comblera cette lacune en centralisant la pléthore de données issues d'évaluations déjà disponibles et en produisant les résumés par pays. Par ailleurs, l'ISU entreprendra des études plus approfondies qui pourraient s'avérer utiles pour les décideurs nationaux, consistant par exemple à suivre et à documenter les nouveaux modèles et tendances en matière d'évaluation, sans évaluer leur qualité ou comparer des résultats d'un pays à l'autre.

Les autorités peuvent ressentir le besoin de réviser les caractéristiques des évaluations auprès des apprenants pour diverses raisons, notamment

- (i) le fait que le coût par tête des évaluations de qualité tend à être élevé. Dès lors, il peut être souhaitable d'examiner l'utilité et la valeur ajoutée de chaque évaluation qui existe ;
- (ii) La confiance du public dans les tests peut être influencée par sa perception de leur équité, ce qui à son tour peut être associé à la qualité du test ; et

- (iii) Les pays pourraient souhaiter aligner leurs évaluations sur les normes internationales. Les résultats du catalogue constituent une ressource utile en vue de soutenir de telles révisions.

Dans le cadre de la stratégie de diffusion des résumés par pays et des rapports d'étude du catalogue de l'ISU, celui-ci prévoit d'offrir à chaque pays l'opportunité de rejoindre un réseau d'échange de pratiques où les autorités sont susceptibles de partager leurs expériences du suivi des résultats de l'apprentissage. L'accent devrait être mis sur les évaluations nationales. Ces futurs réseaux d'échange de pratiques pourraient également tenir lieu de forums de discussion sur les systèmes nationaux et les nouvelles normes internationales.

Le catalogue ne comparera pas les performances des apprenants. Les résultats scolaires des apprenants ne sont pas nécessairement comparables, notamment en ce qui concerne les évaluations nationales, car elles ne sont pas conçues pour être comparables. Néanmoins, la comparabilité ne doit pas être considérée comme un prérequis en vue d'analyser les résultats d'évaluation. Les données non comparables sont utiles pour assurer le suivi des résultats de l'apprentissage, tant à l'échelle nationale qu'à l'échelle internationale.

Quoi qu'il en soit de la comparabilité, les informations relatives aux évaluations des apprenants provenant des bases de données existantes ont prouvé leur utilité en éveillant l'intérêt sur les raisons des différences entre pays. Il est important de souligner que plusieurs facteurs sous-jacents expliquant ces différences pourraient ne pas être évidents à première vue.

Même si les résultats scolaires des apprenants ne sont pas toujours comparables, leurs caractéristiques le sont, qui incluent notamment l'année d'étude, les approches, les domaines évalués, les langues, et la manière dont on présente les résultats et dont on les utilise à des fins de politiques. Les analyses descriptives des attributs des évaluations révèlent des modèles de décision que les autorités prennent afin de faire le suivi des acquis de l'apprentissage. En vue de documenter les modèles et tendances globaux en matière d'évaluation, le catalogue comparera les pays sur la base des attributs descriptifs de leurs évaluations.

Le présent **Manuel** a été élaboré en vue d'aider les fournisseurs de données à compléter le catalogue des mesures des résultats de l'apprentissage. Pour toute information complémentaire relative à la façon de compléter et de soumettre le catalogue, veuillez contacter :

M. Georges Boade par e-mail à l'adresse g.boade@unesco.org; ou par téléphone au 1-514-343-7845.

2. COUVERTURE DU QUESTIONNAIRE

Types d'évaluations à inclure dans l'enquête :

L'enquête se concentre sur les évaluations organisées au niveau du système scolaire et ciblant principalement les apprenants scolarisés dans des programmes d'enseignement formel. Seules les données générées par des évaluations nationales à grande échelle, y compris les examens publics/nationaux (de fin de cycle) administrées aux niveaux 1 et 2 de la CITE, dans l'enseignement général ou technique/professionnel, seront collectées dans cette enquête. Pour toute information au sujet des niveaux de la CITE pour votre pays, veuillez consulter les [cartographies de la CITE 97 ainsi que le cadre de la CITE 2011](#). Les évaluations internationales auxquelles les pays participent doivent être mentionnées dans la partie 2.3, mais les données seront collectées directement auprès des agences à l'origine de ces évaluations; ce qui signifie que les données issues de ces évaluations ne seront pas enregistrées dans ce modèle par les responsables nationaux. Des exemples de telles évaluations internationales sont: LLECE, TIMSS, PASEC et SACMEQ.

Types d'évaluations à exclure de l'enquête :

Trois catégories d'évaluations ne sont pas incluses dans l'enquête :

1. Les évaluations organisées par les écoles et par des institutions d'enseignement décentralisées, et destinées principalement à la sélection en vue d'un passage d'une année d'études à une autre ou d'un programme à un autre. Il s'agit de :

- a. tests organisés généralement au terme de l'enseignement primaire dans le but de sélectionner les apprenants ayant de bonnes performances pour qu'ils intègrent la première année de l'enseignement secondaire général ou des programmes techniques/professionnels;
- b. tests menés auprès des apprenants afin de les sélectionner dans des établissements professionnels ; et
- c. évaluations organisées par des établissements professionnels.

2. Les tests qui ne ciblent pas spécifiquement les apprenants inscrits dans des programmes d'enseignement et auxquels sont soumis tous les candidats à une qualification professionnelle, généralement sans scolarisation préalable dans un programme d'enseignement officiel général ou technique équivalent. Ce type de test est décentralisé et peut être organisé par des syndicats ou des associations professionnelles. Les examens pour l'obtention du permis de conduire ou dans le but de devenir membre d'une association professionnelle en sont deux exemples.

3. Les évaluations des acquis scolaires réalisées au niveau des ménages, même si elles visent des enfants scolarisés dans des programmes de niveau 1 ou 2 de la CITE.

Le questionnaire a été conçu de manière à s'adapter à différents scénarios nationaux. Par exemple :

- ✓ La section 2 peut accueillir 22 titres d'évaluations. Si l'espace réservé dans le questionnaire standard s'avère insuffisant pour un pays donné, veuillez nous en informer pour y ajouter de l'espace supplémentaire.
- ✓ La section 7 : 7.2.3 permet d'encoder les données pour dix sujets/construits obligatoires couverts par une seule évaluation. Si une évaluation couvre plus de 10 sujets/construits obligatoires, veuillez nous en informer pour y ajouter des colonnes supplémentaires.

3. COMMENT COMPLÉTER LE QUESTIONNAIRE

Il est demandé aux répondants de lire attentivement les instructions, définitions et illustrations suivantes en vue de compléter avec précision le questionnaire du catalogue de l'ISU. Le présent manuel fournit toutes les définitions des variables clés utilisées dans le questionnaire et explique le type de données requis pour chaque item ou élément d'un tableau.

Le questionnaire compte huit sections. Chaque section est clôturée par un item nommé « Métadonnées ». Cet espace est prévu pour que les répondants fournissent toute explication ou précision qu'ils jugeraient nécessaire concernant cette section. Cela permettra aussi d'interpréter les données requises.

Si les mêmes informations et données sont requises dans plus d'une question, veuillez-vous assurer que les données saisies sont cohérentes pour toutes les questions. Au cas où une incohérence serait inévitable, veuillez l'expliquer clairement dans le champ Métadonnées à la fin de chaque section.

Veuillez ne pas laisser vides les cellules où une réponse écrite est prévue. L'emploi correct des **codes** correspondant aux données manquantes est essentiel en vue d'assurer l'intégrité des données. En effet, les analyses statistiques et les rapports de l'ISU exigent de distinguer les raisons pour lesquelles les données ne sont pas fournies.

L'ISU encourage les pays à procéder à leur propre évaluation des données manquantes ou incomplètes. Pour signaler une cellule contenant des valeurs estimées, veuillez placer un astérisque (*) devant le nombre, par ex. *68794. Ne laissez pas d'espace entre le symbole et le chiffre. Si, malgré ces efforts, certaines données ne sont pas disponibles ou sont incomplètes, veuillez en expliquer la raison dans une note séparée ou dans le champ Métadonnées à la fin de chaque partie.

Outre les explications ou précisions à noter dans les Métadonnées, veuillez utiliser cet espace afin de citer tout matériel de référence, manuel, publication ou site web auquel vous avez fait référence, ce qui peut aider à comprendre les données de votre pays.

Consultez le **Glossaire** à la fin du présent manuel pour connaître les définitions et explications des données à intégrer dans le questionnaire.

De plus, notez que l'enquête vise à collecter des informations sur des évaluations et les examens publics nationaux administrés entre **2000 et 2014**. Veuillez toujours indiquer l'année des données que vous rapportez.

4. INSTRUCTIONS POUR CHAQUE QUESTION

SECTION 2 : Liste des évaluations

Ne reportez chaque évaluation qu'une seule fois, même si elle couvre plusieurs années d'études ou matières. Considérez divers tests portant le même nom comme une seule évaluation, indépendamment du nombre d'années d'études, matières et langues visés. Tant que le même nom est donné à des évaluations couvrant plusieurs années d'études, il faut les considérer comme faisant partie d'une même évaluation type. Ainsi, si une évaluation est menée la même année dans plus d'une année d'études, ne l'incluez qu'une seule fois dans la liste – cette information complémentaire sera recueillie dans les sections suivantes.

Cette enquête porte sur des évaluations à grande échelle administrées aux élèves des écoles primaires et de l'enseignement secondaire du premier cycle de l'enseignement général ou technique dans votre pays. Elles doivent couvrir la totalité du pays et cibler tous les apprenants, ou un échantillon de ces derniers qui partagent un ou plusieurs attributs tels que l'année d'études dans laquelle ils sont scolarisés, ou l'âge.



*Veillez-vous reporter aux pages précédentes de ce manuel pour en savoir plus sur les **types d'évaluations à prendre en compte dans le cadre de cette enquête**, et au glossaire pour la définition des **évaluations à faibles enjeux et à enjeux élevés** avant de compléter cette partie.*



2.1 Liste des évaluations nationales à grande échelle

Ce tableau vise à obtenir des informations générales sur toutes les évaluations nationales à grande échelle.

Veillez fournir les informations suivantes pour chaque évaluation, dans l'ordre chronologique d'administration, en débutant par l'évaluation la plus récente :

- i) l'acronyme dans sa **langue d'origine** s'il existe (ne jamais traduire les acronymes). Veuillez assigner le code "a" si l'acronyme n'existe pas.
- ii) la dénomination complète de l'évaluation en sa langue d'origine et en français :
 - devant la lettre N, dans sa langue d'origine.
 - devant la lettre F, en français si la langue d'origine ne l'est pas ou si une traduction en français est disponible.
- iii) les programme(s) d'enseignement que l'évaluation couvre. Sélectionnez dans la liste fournie :
 - enseignement primaire seulement (CITE 1)
 - enseignement secondaire de premier cycle seulement (CITE 2)
 - enseignement primaire et secondaire de premier cycle (CITE 1 & 2)
 - enseignement secondaire de premier et de second cycle (CITE 2 & 3)
 - enseignement primaire, secondaire premier et second cycle
- iv) Le nom des institutions responsable de l'administration de cette évaluation/examen
- v) l'enjeu de chaque évaluation listée, qui soit élevé ou faible, pour chacune des populations suivantes: les apprenants, les enseignants et les établissements scolaires. Veuillez prendre note que dans certains cas, l'enjeu d'une évaluation/ examen peut exister de façon informelle. Si c'est le cas, il est important de donner des explications dans la partie des métadonnées.
- vi) l'année (aaaa) la plus récente au cours de laquelle l'évaluation a été administrée.
- vii) au cas où votre pays possède une structure administrative fédérale, sans évaluation nationale couvrant tous les états, provinces ou régions (selon la terminologie utilisée dans votre pays), veuillez sélectionner OUI dans la cellule rouge sous le tableau. Cela signifie que le tableau 2.1 sera vide puisqu'il n'y a pas d'évaluation nationale à grande échelle.

Mentionnez dans l'espace réservée de la sous-section 2.1 toutes les observations/précisions à prendre en compte avec les données fournies afin de donner une image aussi complète et exacte que possible de votre pays.



2.2 Liste des examens publics nationaux

Ce tableau vise à obtenir des informations générales sur les examens publics nationaux. Ici s'appliquent les mêmes instructions que pour 2.1; veuillez donc vous référer au point 2.1 pour des instructions plus détaillées.

Au cas où votre pays possède une structure administrative fédérale, sans examen public national couvrant tous les états, provinces ou régions (selon la terminologie utilisée dans votre pays), veuillez sélectionner OUI dans la cellule rouge sous le tableau. Cela signifie que le tableau 2.2 sera vide puisqu'il n'y a pas d'examen public national.

Mentionnez dans l'espace réservée de la sous-section 2.2 toutes les observations/précisions à prendre en compte avec les données fournies afin de donner une image aussi complète et exacte que possible.



2.3 Liste d'évaluations internationales

Ce tableau vise à obtenir des informations générales sur les évaluations internationales administrées dans votre pays. Cette catégorie inclut des évaluations globales ou toute évaluation administrée dans deux pays ou plus.

Au cas où votre pays possède une structure administrative fédérale et seuls certains états, provinces ou régions (selon la terminologie utilisée dans votre pays) ont participé aux évaluations internationales, veuillez sélectionner OUI dans la cellule rouge sous le tableau. Cela signifie que le tableau 2.3 sera vide puisqu'il n'y a pas eu d'évaluation nationale administrée au niveau national, mais seulement dans un nombre limité d'états, provinces ou régions.

Mentionnez dans l'espace réservée de la sous-section 2.3 toutes les observations/précisions à prendre en compte avec les données fournies afin de donner une image aussi complète et exacte que possible.

Veuillez noter qu'il incombe aux points focaux nationaux de compléter les parties 3 et suivantes relatives aux évaluations nationales et aux examens publics dans leurs pays. Les données relatives aux évaluations internationales seront obtenues directement auprès des institutions responsables du développement de ces évaluations.



2.4 Existence d'évaluations à grande échelle en éducation de la petite enfance (niveau 0 de la CITE)

Cette sous-section vise à mettre en lumière les évaluations à grande échelle de l'enseignement pré-primaire ou toute autre éducation de la petite enfance dans votre pays. Si vous sélectionnez **Oui**, ce qui implique que ce type d'évaluation existe dans les programmes de la petite enfance dans votre pays, veuillez fournir le titre de chaque évaluation, l'âge des enfants qu'elle cible, sa couverture (nationale ou régionale) et les institutions qui sont responsables de son administration.

Mentionnez toutes les observations/précisions susceptibles de compléter les données fournies dans l'espace réservée de la sous-section 2.4.

2.5 Existence d'évaluations des acquis des apprentissages réalisées auprès des ménages

Cette sous-section vise à présenter les évaluations à grande échelle des acquis des apprentissages réalisées auprès des ménages dans votre pays.

Ces évaluations sont généralement menées par des intervieweurs qui rendent visite aux ménages afin de collecter des informations contextuelles sur le ménage et d'évaluer les compétences cognitives d'un ou de plusieurs membres du ménage.

Pour cette raison, il ne faut pas inclure ici les évaluations organisées par les écoles où un enfant ramène un questionnaire à la maison pour que ses parents le complètent. Lorsque c'est le cas, ce questionnaire fait partie des documents en rapport avec les évaluations listées aux sous-sections 2.1 à 2.3.

Les évaluations des acquis des apprentissages réalisées auprès des ménages ne ciblent pas seulement les apprenants scolarisés dans des programmes formels; souvent, elles évaluent également les enfants non scolarisés et peuvent appliquer d'autres critères de sélection, comme l'âge des enfants à évaluer.



À partir de la section 3, vous êtes prié de fournir des informations permettant de détailler chaque évaluation nationale et examen public énumérés dans la partie 2 (sous-sections 2.1 et 2.2 uniquement).

Les données relatives aux évaluations internationales (dans la liste de la sous-section 2.3) seront complétées par l'institution responsable de leur administration.

Avant de commencer à compléter les données concernant ces parties, veuillez faire autant de copies de ce questionnaire qu'il y a d'évaluations énumérées dans les parties 2.1 et 2.2. Afin d'identifier facilement les questionnaires, assurez-vous que chaque copie porte le nom du pays, suivi de l'acronyme de l'évaluation s'il existe (dans le cas contraire, vous pouvez choisir n'importe quel nom permettant d'identifier facilement l'évaluation). Par exemple : Sri Lanka-GCE-O-L ou France-CAP.

Puis, complétez ces parties pour chaque évaluation

SECTION 3: Champ d'application, objectifs, financement et parties prenantes

Cette partie donne des informations qui caractérisent chaque évaluation. Elle comprend des données sur le champ d'application de l'évaluation, ses objectifs, les sources de financement de l'évaluation et les rôles de chaque partie prenante.

Les cellules vertes proposent une liste d'options parmi lesquelles vous devrez en choisir une; il n'est donc pas nécessaire de saisir vos réponses.

Dans certains cas lorsqu'il est requis de spécifier une catégorie ne figurant pas dans la liste fournie « Autre, veuillez spécifier », une cellule jaune est réservée pour y inscrire cette catégorie.

S'il existe des observations qui devraient être prises en compte pour comprendre ces données, veuillez les écrire dans l'espace disponible sous la rubrique 'Métadonnées' à la fin de cette section.

Afin de compléter cette section, il vous est conseillé d'utiliser ou de consulter la dernière version de la documentation officielle et des ressources telles que les cadres d'évaluation, rapports et bases de données.



3.1. Informations générales sur l'évaluation ou l'examen

Veuillez écrire le nom complet de l'évaluation/examen en français et dans la langue nationale ou officielle si ce n'est pas le français, ainsi que l'acronyme, exactement comme dans la section 2. Si l'acronyme n'existe pas, veuillez attribuer le code "a" à l'item 3.1.3



*Veuillez-vous reporter au glossaire pour les définitions des **évaluations à faibles enjeux et à enjeux élevés.***



3.2. Champ d'application de l'évaluation

Il est possible qu'une seule évaluation couvre plusieurs instances de chaque variable. Par exemple, une évaluation peut avoir été conçue en vue de couvrir plusieurs années d'études ou les enfants dans une certaine tranche d'âge. Veuillez sélectionner toutes les catégories concernées par variable.

3.2.1-3.2.2: Pour chacune des variables énumérées ci-dessous, indiquez la ou les catégories qui s'appliquent à cette évaluation:

3.2.1. programme(s) d'enseignement / niveaux de la CITE

3.2.2. année(s) d'études

3.2.3. groupes d'âge

3.2.4. curriculum scolaire ciblé. Si Autre, veuillez spécifier dans la cellule jaune, et inscrire toutes les explications dans l'espace réservée aux métadonnées à la fin de section

3.2.5. types d'établissement scolaire. Si Autre, veuillez spécifier dans la cellule jaune.

3.2.6. Veuillez sélectionner l'orientation du programme (technique/professionnel, général ou les deux). L'évaluation cible habituellement un programme d'enseignement général ou professionnel/technique ; néanmoins, au cas où l'évaluation couvrirait les deux, veuillez écrire toutes les explications nécessaires en notes dans l'espace réservée aux métadonnées à la fin de la section.

3.2.7. Veuillez sélectionner le format de l'évaluation (ou du test) (écrit, oral, pratique ou une combinaison). Concernant les évaluations/examens exigeant des apprenants qu'ils passent des tests par écrit, sélectionnez Oui pour «écrit» ; lorsque les compétences des apprenants sont déterminées uniquement à l'aide de tests oraux, sélectionnez « oral ». Si le test inclue aussi la pratique, veuillez sélectionner Oui pour «pratique». Si cette évaluation/examen combine les trois formats écrit, oral et pratique, veuillez fournir toutes les explications nécessaires dans les métadonnées à la fin de la section.



*Veuillez-vous reporter au glossaire pour les définitions d'**année d'études, niveaux de la CITE et construit.***



3.3. Objectifs de l'évaluation

Veillez sélectionner OUI à toutes les propositions qui s'appliquent et ajoutez dans la cellule jaune tout objectif supplémentaire qui ne serait pas inclus dans la liste prédéfinie.



3.4. Sources de financement

Veillez sélectionner OUI à toutes les source(s) de financement qui s'applique(nt) à cette évaluation. Dans la cellule jaune, ajoutez toute source de financement supplémentaire qui ne figure pas dans la liste prédéfinie. Il est possible qu'il y ait plusieurs sources de financement provenant de plusieurs parties prenantes. Veuillez fournir les données désagrégées si elles sont disponibles dans l'espace réservée aux métadonnées à la fin de la section. Un exemple de donnée désagrégée serait le pourcentage du financement total par un bailleur de fonds.

Note: Pour les besoins de cette enquête, les organisations étrangères sont celles qui sont créés, gérées et financées majoritairement par les citoyens et résidents de pays autres que le vôtre, tandis que les organisations locales le sont par les citoyens et résidents de votre pays. Les organisations internationales se réfèrent à celles qui sont créées, gérées et financées par plus d'un pays. Pour cette raison, les bureaux nationaux des organisations internationales comme la Banque Mondiale, la FAO, l'UNICEF ou l'UNESCO sont toujours considérés dans la catégorie des organisations internationales même si l'essentiel de leur financement est en majorité de source nationale.



3.5. Les parties prenantes et leurs rôles

Nous proposons une liste indicative avec les parties prenantes potentielles dans les lignes et leurs rôles possibles dans les colonnes. Veuillez sélectionner l'option OUI dans les combinaisons partie prenante/rôle applicables. Une partie prenante peut jouer plusieurs rôles. En outre, plusieurs parties prenantes peuvent partager un rôle donné. Veuillez sélectionner OUI à toutes les combinaisons qui s'appliquent et ajouter toute autre partie prenante impliquée dans l'évaluation/examen dans la cellule jaune de la dernière ligne, et tout autre rôle qui n'aurait pas été mentionné dans la cellule jaune à la dernière colonne.



Dans les pays qui possèdent une **structure administrative fédérale**, les situations peuvent être très variables d'un pays à l'autre et d'une région du pays à une autre. Dans ce cas, veuillez inscrire dans l'espace Métadonnées toutes les informations nécessaires pour comprendre et documenter cette situation à l'attention des utilisateurs des données.



4.1. Conception du test

4.1.1. Tous les candidats reçoivent le même livret de tâches ou les mêmes épreuves lorsqu'ils sont évalués sur le même contenu.

Toutefois, il est aussi possible que les apprenants participant à une évaluation donnée soient testés sur des contenus différents ; cela peut être le cas lorsque trop d'items doivent être administrés. Étant donné les contraintes de temps et le fardeau sur le répondant, seule une partie de la batterie d'items est administrée à chaque apprenant. Les apprenants reçoivent donc différents livrets de tâches avec une sélection d'items. Les livrets de tâches sont composés selon une méthode connue sous le nom d'échantillonnage matriciel complexe, mais avec un ensemble d'items communs à tous les livrets.

Il existe aussi des situations dans lesquelles les contenus des livrets sont fonction des objectifs du programme à évaluer de sorte que chacun d'eux couvre un sous-ensemble de ces objectifs, et sont administrés à des groupes différents sans qu'il y ait des tâches communes à tous les candidats évalués.

Veillez indiquer la méthode de sélection de l'évaluation. Veuillez-vous reporter au glossaire pour la définition d'échantillonnage matriciel complexe.

4.1.2. Dans ce tableau, on vous demande d'indiquer les modes d'administration du test et de collecte des données utilisés pour mener à bien l'évaluation. Pour chaque mode d'administration pertinent, veuillez sélectionner le(s) mode(s) de collecte de données approprié(s) :

- i. un mode de collecte de données en face à face où les apprenants passent le test en présence de superviseurs (personnel d'évaluation).
- ii. un mode de collecte de données en ligne où les apprenants passent le test via internet sans supervision par une équipe d'évaluation. Notez que le temps alloué pour terminer l'évaluation peut varier.
- iii. un mode de collecte de données par la poste où les apprenants reçoivent le test par la poste et le complètent sans supervision par une équipe d'évaluation.
- iv. si un mode d'administration ou un mode de collecte de données n'est pas présent dans la liste, veuillez l'inscrire dans la cellule jaune de dernière rangée ou colonne respectivement et donner davantage d'informations si nécessaire dans l'espace des métadonnées à la fin de la section.



*Veillez-vous reporter au glossaire pour la définition de **test fixe informatisé et test adaptatif informatisé**.*

4.1.3. Il pourrait y avoir un mécanisme mis en place pour faire une analyse de l'évaluation avant ou après son administration. Par exemple, avant l'administration, il est possible d'analyser l'évaluation pour s'assurer que les outils d'évaluation correspondent au programme d'études ou aux domaines à observer. Ou encore, après l'évaluation et l'analyse des résultats, on peut les examiner afin d'en tirer des leçons et adapter les futures évaluations. Veuillez spécifier si de telles analyses sont réalisées. Incluez tout autre mécanisme mis en pratique dans la cellule jaune, et donnez les détails dans l'espace des métadonnées.



4.2. Caractéristiques des items/questions, et questionnaires contextuels

L'information requise pour cette sous-section peut être collectée à partir d'un échantillon des carnets de test de la plus récente administration de l'examen ou de l'évaluation.

4.2.1. Format des items/questions

Veillez sélectionner OUI au format approprié :

- i. Deux options : les cas d'items ou de questions dont la réponse doit être choisie entre deux options prédéfinies seulement.
- ii. Trois ou plus d'options : les cas d'items ou de questions dont la réponse doit être choisie dans une liste de trois options prédéfinies ou davantage.
- iii. Réponse ouverte brève : les cas d'items ou de questions où les apprenants reçoivent un score (avec une fourchette allant d'un minimum à un maximum) sur la base d'une brève réponse écrite dans leurs propres mots.

- iv. Rédaction si les apprenants doivent élaborer une réponse écrite structurée sous la forme de plusieurs paragraphes.
- v. Projet individuel : dans le cadre d'évaluations techniques/professionnelles, l'application de connaissances ou de compétences au travers d'un projet pratique spécifique fait généralement partie de l'évaluation. Veuillez sélectionner Oui à cette option si les apprenants doivent faire une présentation devant un jury ou une équipe d'examineurs concernant leur projet, ou s'ils doivent soumettre une production donnée afin de démontrer qu'ils sont capables d'exploiter une certaine technique.
- vi. Projet de groupe : si le projet doit être mené par un groupe d'apprenants.
- vii. Autre : veuillez inscrire dans la cellule jaune tout autre format utilisé dans le cadre de l'évaluation et qui n'est pas mentionné parmi les formats de questions ou items fournis.

4.2.2. Format du stimulus

Le stimulus donne le contexte auquel les apprenants se réfèrent pour répondre aux questions ou réaliser certaines tâches.

Les stimuli peuvent avoir plusieurs formats. Veuillez sélectionner OUI à tous les formats qui s'appliquent :

- i. Texte : si du texte est utilisé en vue de communiquer le contexte à l'apprenant, quel que soit le support utilisé. Par exemple, le texte peut être fourni sur papier, ou présenté au travers d'un écran. Le texte peut également prendre plusieurs formes y compris des textes continus/suivis ou non tels que des listes, des tableaux, des horaires, des cartes, etc. Veuillez-vous reporter au glossaire pour des définitions de textes continus ou non continus, mixtes et multiples.
- ii. Audio : si le stimulus est uniquement sonore. Par exemple, les personnes évaluées peuvent être priées d'écouter une histoire courte et de répondre à quelques questions, sans disposer du texte écrit correspondant. Ce contenu peut provenir d'un enregistrement ou être lu par une personne.
- iii. Vidéo : si le stimulus utilisé est une combinaison de sons et d'images.
- iv. Autres : Veuillez inscrire dans la cellule jaune, tout autre format inclus dans cette évaluation mais qui n'est pas mentionné parmi les formats de stimuli listés. Donner aussi plus de détails dans les métadonnées.



Veuillez-vous reporter au glossaire pour la définition de texte continu, non continu, mixed ou multiple.

4.2.3 - 4.2.4.

Apprenants ayant des besoins spéciaux : Veuillez-vous reporter au glossaire pour la définition de l'autisme, de la déficience auditive, de la surdi-cécité, de la surdit , des troubles  motionnels, de la d ficience visuelle et des troubles de la parole.

4.2.4 (i) Veuillez sélectionner OUI pour le type de handicap pour lequel l' valuation est adapt e.

4.2.4 (ii) Veuillez d crire les types d'adaptation(s) fournis, par type de handicap. Par exemple, si l' valuation est destin e aux personnes atteintes de d ficience visuelle, elle peut  tre disponible en braille, ou les stimuli peuvent  tre pr sent s dans un format diff rent – sonore au lieu de textuel, par exemple.

Veuillez noter que si l'option choisie au 4.2.4.i est "Non", les cellules appropri es en 4.2.4.ii se remplissent automatiquement avec le code "a" pour non applicable.

4.2.5. Si les personnes  valu es poss dent une ma trise limit e de la langue d' valuation, veuillez d crire le type d'adaptation(s) dans la cellule appropri e. Par exemple, les personnes  valu es peuvent  tre autoris es   donner leurs r ponses   une  valuation dans une autre langue. Cela peut  tre courant dans des environnements bilingues. Ils peuvent  galement recevoir du temps additionnel pour leur permettre de terminer l' valuation.

4.2.6. Questionnaires contextuels

Dans ce tableau, on vous demande de fournir des informations g n rales concernant les diff rents questionnaires contextuels qui forment l' valuation g n rale. Pour chaque questionnaire :

-  crire son nom dans la premi re colonne
- s lectionner le groupe auquel il incombe de le compl ter. S'il n'est pas dans la liste, veuillez le fournir dans les m tadonn es en fin de section au bas de la feuille.
- donner plus de d tails sur la dur e officielle pour le remplir (si vous utilisez des estimations, assurez-vous de placer un ast risque devant le chiffre)
- d tailler l'objectif poursuivi par le questionnaire.
- donner la nature des donn es collect es. Les donn es d mographiques, le m nage et la structure de la famille, l' ducation, la zone de r sidence figurent parmi les groupes de variables auxquels on recourt couramment afin de collecter des donn es contextuelles.

Exemple : Dans PISA 2012, le principal test d'évaluation comprend des épreuves en mathématiques, la compréhension de l'écrit, les sciences, la résolution de problèmes, la culture financière et les TIC. En outre, le programme comprend des questionnaires contextuels qui doivent être complétés par des apprenants, des chefs d'établissement et des parents et qui fournissent des informations contextuelles. Cette information est généralement utilisée à des fins d'analyse.



Veillez envoyer à l'ISU les dernières copies numériques des questionnaires contextuels en même temps que le questionnaire dûment rempli pour chacune des évaluations traitées dans cette enquête. Notez que ces questionnaires seront rendus publics, sauf mention contraire.

4.2.7. Cadres d'évaluation

Dans plusieurs évaluations nationales, les outils de collecte de données sont élaborés en se fondant sur des cadres conceptuels spécifiques développés par l'équipe nationale d'évaluation. Ces cadres précisent la raison d'être de la structure de l'évaluation. Par exemple, les cadres d'évaluation de PISA 2012 sont : cadre d'évaluation de la culture mathématique, cadre d'évaluation de la compréhension de l'écrit, cadre d'évaluation de la culture scientifique, cadre d'évaluation de la résolution de problèmes et cadre d'évaluation de la culture financière.

4.2.8. Si elle est disponible, veuillez fournir la liste des construits supposés être mesurés dans l'évaluation. Généralement, elles figurent dans les cadres d'évaluation.



Veillez envoyer les dernières copies numériques des cadres d'évaluation ou des listes de construits couverts par l'évaluation (si elles existent) à l'ISU. Notez que ces questionnaires seront rendus publics, sauf mention contraire.



5.1. Couverture

Cette enquête touche aussi à la couverture géographique et démographique de l'évaluation.

5.1.1. Couverture géographique : Veuillez sélectionner dans la liste prédéfinie la zone géographique que cette évaluation couvre. Notez que cette enquête vise les évaluations et les examens publics ayant couverture nationale.

5.1.2. Population étudiée : Veuillez décrire et caractériser au mieux la population cible de l'évaluation en vue de permettre la meilleure compréhension possible des personnes évaluées. Par exemple, la population cible peut être décrite en termes de niveaux d'enseignement ou d'année d'études, d'âge, de zones de résidence (campagne, ville), de sexe, de niveau de scolarisation, de propriété des établissements scolaires (publics et non publics), de besoins spéciaux, etc.

5.1.3. Enfants non scolarisés

i) Veuillez sélectionner **Oui** si des enfants non scolarisés sont inclus dans l'évaluation, et allez au 5.1.3 (ii). Sinon, allez à l'item 5.1.4 puisque les cellules en 5.1.3.ii et 5.1.3.iii se remplissent alors automatiquement avec le code "a" pour non applicable.

ii) Veuillez sélectionner **Oui** si les enfants non scolarisés participants fréquentent des programmes d'alphabétisation ou de seconde chance de niveau équivalent qui couvrent le programme d'études ciblé par l'évaluation (si et quand cela est applicable). Par exemple, cela peut être le cas si ces enfants sont échantillonnés dans des programmes de littératie.

iii) Veuillez sélectionner **Oui** si la réussite à l'évaluation permet aux enfants non scolarisés de réintégrer le système scolaire dans l'année d'études suivante appropriée.

Veillez noter que certains examens publics nationaux sont administrés à tous les candidats qu'ils soient scolarisés ou non. Réussir à de tels examens peut être une condition pour accéder dans des programmes scolaires spécifiques, peu importe la trajectoire scolaire suivie.

5.1.4. Certains groupes peuvent être exclus de l'évaluation pour diverses raisons liées ou non à l'objectif de l'évaluation. Par exemple, les apprenants atteints de déficience visuelle peuvent être exclus parce que la conception de l'évaluation ne prévoit pas de modalités spécifiques permettant leur participation.

Le cas échéant, veuillez décrire quels groupes spécifiques sont officiellement exclus de l'évaluation.

D'autres enfants peuvent être exclus car ils résident dans des zones difficiles d'accès. Veuillez décrire les groupes exclus de l'évaluation.

5.1.5. Veuillez indiquer les lieux où les apprenants sont invités à passer l'évaluation, qu'il s'agisse de villes importantes uniquement ou de centres locaux disséminés dans le pays (y compris des établissements d'enseignement dans des villages, dans de petites ou grandes villes, etc.). Pour tout autre lieu non listé où les apprenants passent le test, veuillez spécifier dans la cellule jaune.



5.2. Participation

Lorsque les évaluations nationales ne ciblent pas la totalité de la population d'apprenants à un niveau, une année d'études ou un âge donné, il est possible d'évaluer un **échantillon** représentatif de cette population. Pour construire un échantillon représentatif, on se fonde sur plusieurs caractéristiques ou variables comme le nombre d'établissements scolaires et leur répartition entre zones urbaines et rurales et grandes villes, le nombre de classes, le nombre d'apprenants candidats par école/classe, l'âge et le sexe des candidats. Ces informations sont généralement disponibles dans le rapport technique de l'évaluation, et elles sont utilisées à posteriori en vue de valider, encoder et ajuster les données avant de faire des analyses.

Veillez compléter à la fois la population cible totale d'apprenants (N) et le nombre total de personnes évaluées (n) par année d'études couverte par l'évaluation/examen et par sexe.

✓ Les évaluations peuvent couvrir une année d'études ou davantage. Veuillez sélectionner la ou les années d'études couvertes (comme dans 3.2.2). Une année d'études peut être sélectionnée par rangée.

- ✓ Dans le cas d'une évaluation basée sur un échantillon, le nombre total de personnes évaluées (n) correspondra à la taille de l'échantillon et le total (N) sera le nombre d'apprenants identifié dans la base d'échantillonnage de la population d'apprenants cible.
- ✓ Dans le cas d'une évaluation sur la base d'un recensement, le total (N) correspondra au nombre d'apprenants inscrits pour passer le test, alors que le total des personnes évaluées (n) est le nombre d'apprenants qui ont effectivement passé le test. Dans ce cas, n et N devraient être égaux ou presque.

Dans la dernière colonne, il est requis de fournir l'âge des candidats. Si possible, veuillez différencier les âges visés aux niveaux scolaires dont les programmes sont couverts par cette évaluation/examen, des autres tranches d'âges des personnes à qui l'évaluation/examen est administré(e). Par exemple, vous pourrez écrire "10-15; 8-9 et 16+", et dans les métadonnées, vous pourrez spécifier que le test vise les candidats âgés entre 10-15 ans, mais étant donné qu'il est ouvert à tous, on trouve aussi des candidats âgés de 8 à 9 ans et de 16 ans et plus.

Veillez noter que si l'option 'Recensement' a été choisie en 3.1.4, alors toutes les cellules en 5.3 et 5.4 se remplissent automatiquement avec le code "a" pour non applicable. Dans ce cas, continuez directement à la section.

5.3. Échantillonnage

5.3.1. Plusieurs méthodes d'échantillonnage peuvent être utilisées afin de construire un échantillon représentatif d'une population cible. Toutefois, elles dépendent, notamment, de la structure de la base de sondage, du budget, du niveau de ventilation désiré des données et de la structure des rapports à présenter aux décideurs politiques. Tout particulièrement, dans les enquêtes sur l'éducation, les échantillons sont constitués à différents niveaux, partant des unités administratives aux apprenants eux-mêmes. Les plans d'échantillonnage les plus courants sont fournis sous forme d'options prédéfinies dans la question 5.3.1. Aux fins de l'enquête, veuillez sélectionner le plan d'échantillonnage utilisé pour former les échantillons des unités administratives (régions/provinces/états/départements/arrondissements/districts, etc.), des établissements scolaires, des classes et des apprenants (si applicable).

5.3.2. Veuillez fournir un bref résumé descriptif de chaque base de sondage utilisée pour constituer le ou les échantillons.

5.3.3. Veuillez fournir un bref résumé de l'exclusion si cela s'applique. L'exclusion est une source d'erreur, connue sous le nom de sous-couverture. Cela se produit lorsque des bases de sondage inexactes, incomplètes, inadéquates ou dépassées sont utilisées afin de constituer l'échantillon.

5.4. Taux de participation et ajustements

Veillez-vous reporter au glossaire pour les définitions des taux de participation et des méthodes d'ajustement avant de compléter ce tableau.



Veillez-vous reporter au glossaire pour la définition des termes *échantillon*, *méthode d'échantillonnage*, *taux de participation* et *taux de participation pondéré*.



6.1. Traitement des données

6.1.1. Vérification des données: cette étape se situe entre la collecte des données et leur saisie. Elle est nécessaire afin d'assurer que les données sont clairement indiquées sur le questionnaire ou le livret de tâches avant d'être encodées. Veuillez sélectionner OUI à la méthode utilisée pour éditer les données collectées avant qu'elles ne soient encodées. Lorsque la formule « papier-crayon » est employée pour collecter les données, il peut arriver que la **vérification des données** soit réalisée directement sur papier avant que celles-ci soient encodées dans la base de données. Il y a aussi des cas où les données sont directement encodées dans une base de données, mais ce n'est pas un programme informatique qui réalise la vérification, qui reste alors basée sur un contrôle visuel des données. Dans les deux cas, veuillez choisir OUI au **contrôle manuel ou visuel**. Si on recourt à un programme informatique afin de mener à bien ce contrôle, quelle que soit la manière dont les données ont été encodées dans l'ordinateur, veuillez choisir l'option OUI sous '**automatique**'. Si les deux méthodes sont utilisées au cours du processus de vérification des données, veuillez sélectionner les deux. Dans le cas où d'autres méthodes de vérification des données ont été utilisées, veuillez spécifier la méthode dans la cellule jaune à la dernière colonne.

6.1.2. Saisie des données :

En général, que l'évaluation soit informatisée ou de type « papier-crayon », les données doivent être centralisées dans une base de données en vue d'en faciliter la gestion (nettoyage, recherche, analyse, partage, etc.). Plus spécifiquement dans le cas des épreuves de type « papier-crayon », les données disponibles sur papier doivent être encodées avec soin dans la base de données, soit manuellement par un opérateur spécialisé soit au travers d'un dispositif de scanning. Dans ce dernier cas, la numérisation transforme les documents papier en images ou données électroniques susceptibles d'être exploitées par des applications informatiques et des archives. Depuis peu, on fait plus souvent appel à la numérisation en vue d'accélérer le traitement des données provenant de recensements généraux de la population et de l'habitat, ou d'enquêtes à grande échelle auprès des ménages, que la collecte s'effectue en mode face à face ou par courrier. Veuillez indiquer si la saisie des données est manuelle, par scanning ou autre (et spécifier dans la cellule jaune).

6.1.3. Lieux d'édition et de saisie des données :

Selon la culture organisationnelle pour chaque évaluation, la vérification et la saisie des données peuvent être réalisées dans des centres d'évaluation locaux disponibles dans le pays. Par exemple, il peut s'agir des lieux où les apprenants se sont rendus pour passer le test, dans des centres régionaux ou au siège de l'institution qui administre l'évaluation ou est mandatée en vue de mener à bien ces activités. Veuillez sélectionner OUI à chacun des lieux qui s'appliquent à cette évaluation.



*Veuillez-vous reporter au glossaire pour la définition de **traitement des données, édition des données, saisie des données et codage des données**.*



6.2. Évaluation des données

Les données nécessaires pour cette sous-section sont généralement disponibles pour les utilisateurs dans les rapports techniques.

6.2.1. & 6.2.2. Veuillez vérifier si un rapport technique de cette évaluation/examen est disponible pour cette évaluation. Il vous sera plus facile de compléter cette partie avec les données appropriées relatives aux estimations des erreurs d'échantillonnage et aux actions entreprises afin de réduire les erreurs non dues à l'échantillonnage. Si un rapport technique est disponible en ligne, veuillez fournir son URL dans les métadonnées à la fin de la section.

6.2.3. Les candidats sont classés dans la catégorie non-réponse si:

- i. aucune donnée n'a été collectée parce qu'il a été impossible de les joindre ou ils ont préféré ne pas prendre part à l'évaluation. Les données sont totalement manquantes.
- ii. le nombre de questions/items/tâches achevés n'est pas suffisant (inférieur à un minimum à spécifier, en pourcentage du nombre total de questions/items/tâches) pour prendre en compte leurs données à des fins d'analyse plus poussée.

Veillez écrire le taux standard pour cette évaluation si le pourcentage d'achèvement est la règle afin d'identifier les non-répondants. Si une autre règle s'applique, veuillez l'inscrire dans la cellule jaune et l'expliquer dans les métadonnées en fin de section au bas de la feuille.

6.2.4. Non réponses

Veillez indiquer les critères appliqués dans cette évaluation en vue de classer des items dans la catégorie non-réponse (donnée manquante). Si un autre critère non listé est utilisé, veuillez l'inscrire dans la cellule jaune et si possible, donner des explications dans l'espace réservé aux métadonnées.



Ajoutez toute observation/précision importante en vue de comprendre et d'interpréter les données dans l'espace Métadonnées à la fin de la partie.

SECTION 7: Mesures et résultats

Veillez-vous reporter aux définitions et exemples donnés dans le manuel avant de compléter cette partie.



7.1. Mesure

7.1.1. Catégories de modèle

i) Les définitions des modèles de la théorie classique des tests (TCT), théorie de la réponse à l'item (TRI) et de Rasch se trouvent dans le glossaire. Des informations sur l'utilisation de ces modèles sont généralement disponibles dans le rapport technique de l'évaluation et peuvent également être ajoutées aux autres rapports où l'on présente les résultats des apprenants. Veuillez spécifier toutes les théories de mesure qui s'appliquent à cette évaluation, et préciser dans la cellule jaune toute autre théorie à laquelle on a eu recours mais non énumérée dans la liste, en vue de mesurer les résultats des apprenants.

ii) Si TRI est sélectionnée ci-dessus, veuillez spécifier le modèle utilisé, qu'il s'agisse du modèle Logistique, de Rasch ou toute autre catégorie de modèles non cités. Si l'option sur la TRI en 7.1.i est "Non", alors les cellules en 7.1.1.ii se remplissent automatiquement avec le code "a" pour Non applicable

iii) Choisissez dans la liste le mode de présentation des résultats : par contenu du domaine évalué, globalement ou les deux. La présentation par domaine signifie que les résultats sont disponibles pour chaque contenu du domaine de connaissance théorique ou sous-domaines qui, ensembles forment le domaine évalué. Par exemple, une évaluation des mathématiques peut comprendre trois sous-domaines différents : algèbre, géométrie et statistiques.

7.1.2. Sur la base des rapports analytiques ou techniques, il est possible d'identifier les méthodes utilisées pour le référencement des performances des candidats. Veuillez spécifier toutes celles qui s'appliquent dans cette évaluation et indiquer dans la cellule jaune toute autre méthode de référencement utilisé dans cette évaluation/examen, mais non listée.



Veillez-vous reporter au glossaire pour la définition de : **théorie classique des tests (TCT), théorie de réponse à l'item (TRI), modèle de Rasch, référencement normatif, référencement critique, référencement individuel.**



7.2. Résultats

7.2.1. Cette sous-section ne doit être complétée **que si l'examen /évaluation décrit(e) dans la partie 3.1 n'a pas de filières et/ou de domaines de spécialisation.** Sinon, veuillez passer les sous-sections 7.2.1 à 7.2.3 et aller à 7.2.4.

Veillez sélectionner le type de mesure qui s'applique à l'évaluation, et donner la valeur minimale nécessaire pour qu'un apprenant atteigne la norme nationale. **Veillez** noter que si l'option choisie est "Non", les cellules appropriées correspondant au minimum requis du standard national se remplissent automatiquement avec le code "a" pour non applicable.

Par exemple, si le pourcentage d'items réussis est utilisé, on déclare que les apprenants atteignent la norme nationale minimale s'ils répondent correctement à au moins 50 % des items, pour autant que tous les items aient la même pondération. Dans ce cas, la norme nationale minimale est 50 %.

Généralement, quand le score (moyen) sur l'échelle de maîtrise est utilisé pour rapporter les performances des apprenants, la norme nationale minimale est identifiée par un nombre brut. Dans l'hypothèse où l'exigence minimale est 500, on considère que l'apprenant qui obtient 500 ou plus a atteint la norme nationale.

Actuellement, on fait de plus en plus souvent appel aux niveaux de compétences en vue de refléter et de classer les différences dans les résultats des apprenants. Classer les apprenants par groupes permet une intervention fondée sur des besoins ciblés. Cela facilite la mise en œuvre de politiques adaptées ou le développement de programmes d'instruction qui viennent soutenir des groupes d'apprenants ayant des besoins spécifiques. On emploie souvent les lettres de l'alphabet A, B, C, D, etc. ou des séries de nombres. Quoi qu'il en soit, la relation entre ces deux formats est fixe. Chaque lettre ou série décrit les types de compétences que les apprenants maîtrisent ou non. Si cela s'applique à cette évaluation, veuillez sélectionner OUI à la case correspondant aux niveaux de compétences et donnez des explications complètes ou des descripteurs de chaque niveau tels qu'indiqués dans la dernière version des documents officiels de cette évaluation. Veuillez-vous assurer de bien compléter les descriptions des niveaux de compétence dans l'espace Métadonnées à la fin de la section.

Notez qu'en ce qui concerne les évaluations internationales, la métrique et la norme minimale doivent être celles définies par l'institution qui réalise le test, et non celles choisies par les pays participants.

7.2.2. Dans certains pays, les évaluations réalisées à l'école (tests développés et gérés au niveau de l'école) ont un certain poids dans le score total final, notamment pour ce qui est des examens publics. Si c'est le cas pour l'évaluation concernée, veuillez écrire cette information dans la cellule appropriée, et donner des explications complémentaires dans les Métadonnées. Sinon (les évaluations réalisées à l'échelle de l'école ne sont pas reprises dans cette enquête), écrire la lettre 'a' pour 'sans objet'.

7.2.3: Pourcentage des apprenants ayant atteint une performance supérieure à la norme nationale minimale requise selon le sexe et l'âge

- i. Certaines évaluations/examens publics à grande échelle sont menées dans plus d'une année d'études, ce qui explique pourquoi certaines matières peuvent être évaluées spécifiquement dans une année d'étude, ou dans plusieurs années d'études. Le tableau offre de l'espace pour quatre années d'études. Si l'évaluation couvre plus que quatre années d'études, vous pouvez ajouter des items similaires à ceux de 7.2.3.i (copier les lignes entières et les insérer sous 7.2.3.iv).
- ii. Si l'évaluation ou l'examen public couvre un niveau seulement, veuillez choisir l'option "a" pour non applicable du point 7.2.3.ii au 7.2.3.iv. Le cas échéant, toutes les cellules de ces tables seront automatiquement rempli du code "a" pour non applicable.
- iii. D'abord, veuillez écrire chaque discipline/construit ou domaine du programme d'études obligatoire qui est évalué dans les cellules jaunes de la deuxième rangée. Par exemple, les mathématiques, l'anglais, l'économie, l'agriculture, la géographie et la comptabilité sont quelques-unes des matières évaluées dans le cadre des examens publics nationaux.

S'ils sont disponibles, veuillez inscrire les différents construits ou domaines mesurés. Alors que les matières sont enseignées dans un module spécifique et dans un cadre d'apprentissage standard, les construits sont associés à des compétences ou aptitudes spécifiques qui apparaissent en tant que résultat de la combinaison de connaissances acquises au travers de plusieurs disciplines, de l'expérience personnelle et du contexte. Par exemple, la numératie n'est pas une discipline enseignée par elle-même, mais une caractéristique mesurée dans plusieurs évaluations. On peut également citer les différentes littératies rencontrées dans la littérature relative aux compétences du 21^e siècle comme exemples de construits.

Par exemple, l'enquête [PISA 2015](#) évalue quatre construits chez les apprenants de 15 ans : *culture mathématique, compréhension de l'écrit, culture scientifique et compétences et aptitudes à la résolution de problèmes selon une approche collaborative*.

Pour l'ensemble du test et par matière/construit testé, veuillez indiquer le pourcentage d'apprenants dont les performances sont égales ou supérieures à la norme nationale minimale requise pour les deux sexes, pour les hommes uniquement et pour les femmes uniquement. En outre, veuillez indiquer les pourcentages d'apprenants dont les performances sont égales

ou supérieures à la norme nationale minimale requise pour chaque âge des personnes évaluées. Veuillez mettre les âges appropriés entre parenthèses.

Exemple : Dans le pays X, une évaluation nationale à grande échelle cible les apprenants de l'année d'études 6. Les âges correspondants des personnes évaluées tels qu'ils sont enregistrés dans la base de données de l'évaluation sont 10, 11, 12, 13, 14, 16, 17, 20 et 23.

Officiellement, seuls les apprenants de moins de 14 ans peuvent être scolarisés dans l'année d'études 6 dans ce pays. Toutefois, il s'agit d'un examen public et donc même les candidats non scolarisés peuvent participer au test, du moment qu'ils sont qualifiés et qu'ils sont inscrits. Dans ce cas spécifique, il est pertinent d'indiquer les données séparément pour les âges 10, 11, 12, 13 et 14, et pour 15 ans et plus. Si vous utilisez cette fourchette d'âge, veuillez donner des explications dans les Métadonnées qui informeront l'utilisateur de ces données.

Dans chacune des tables, vous devez aussi indiquer la langue en laquelle ces disciplines ou construits **obligatoires** sont testés, leur durée officielle ainsi que le poids de chacun des domaines couverts dans chacune des disciplines ou chacun des construits en termes de pourcentage du score total dans la discipline/construit. Par exemple, si l'on considère que les mathématiques incluent quatre domaines dont l'algèbre, la géométrie, la probabilité et les statistiques; l'algèbre peut compter pour 25% du score total en mathématiques, la géométrie pour 30%, la probabilité pour 15% et les statistiques pour 30%.

Le cas échéant, veuillez aussi reporter le poids de chacun des sujets ou construits **obligatoires** en termes de pourcentage du score total moyen du candidat. Ceci arrive souvent dans le cadre des examens publics nationaux. Par exemple en Mauritanie, le CEPAS/CEF est un examen public national organisé chaque année par le Ministère de l'Éducation Nationale. Il est administré sur base volontaire à tous les élèves en fin du cycle (6e année) primaire (CITE 1). Il comprend sept matières/disciplines obligatoires ci-dessous, et la langue du test est l'Arabe ou le Français. Les titres des disciplines, le temps alloué en minutes, la langue de l'épreuve ainsi que les poids respectifs en termes de pourcentage du score total moyen à l'examen pour chacune des disciplines sont ci-dessous résumés:

Discipline/sujet	Temps alloué (mn)	Langue de l'épreuve	Poids (% score total moyen)
Arabe	90	Arabe	25
Mathématiques	90	Arabe	25
Français	60	Français	15
Histoire et géographie	45	Arabe	10
Sciences naturelles	45	Français	10
Réligion	45	Arabe	10
Éducation civique	30	Arabe	5

Les résultats des évaluations peuvent aussi être reportés par discipline ou au niveau de chacun des domaines de la discipline couverts dans le sujet, sans qu'il soit utile d'agréger le score. C'est le cas dans certains examens publics nationaux lorsque les élèves doivent être évalués sur des disciplines obligatoires et d'autres qui sont optionnelles afin d'obtenir leur certification. Le cas échéant, veuillez reporter seulement le poids par domaines du construit couverts en termes de pourcentage du score total dans la discipline/construit, si cette information est disponible, tandis que la dernière ligne du tableau doit être remplie avec le code "a" pour non applicable.

7.2.4 : Cette sous-section ne doit être complétée que si l'évaluation décrite dans la partie 3.1 a des filières et/ou des domaines de spécialisation (à compléter dans les cellules).

- i. D'abord, veuillez inscrire chaque filière dans les cellules appropriées.
- ii. Ensuite, pour chaque filière, choisissez la métrique utilisée et indiquez le minimum requis pour atteindre la norme nationale, la pondération des évaluations réalisées par les écoles et la liste des domaines de spécialisation disponibles dans la filière.

7.2.5: Les filières fournies ci-dessus apparaissent automatiquement dans la première rangée de ce tableau.

Ne complétez ce tableau que si les filières n'ont pas de domaine de spécialisation.

Pour chaque filière (mais non les matières testées dans la filière), veuillez indiquer les résultats agrégés pour les deux sexes, les hommes uniquement, les femmes uniquement et selon les âges.

Dans la dernière rangée de ce tableau, veuillez inscrire la liste des matières obligatoires testées par filière.

7.2.6: Utilisations des résultats

Les utilisations les plus fréquentes des résultats d'évaluations sont énumérées dans le tableau 7.2.6. Veuillez indiquer à quelles fins les résultats de cette évaluation ont déjà été utilisés, et ajoutez toute information supplémentaire dans la partie réservée aux métadonnées. C'est très important car une partie des rapports par pays est consacrée à l'impact de l'évaluation sur l'élaboration des politiques. L'utilisation est différente de l'objectif : l'objectif d'une évaluation correspond aux finalités pour lesquelles elle a été conçue, qui sont en général clairement énoncées par les promoteurs de l'évaluation à l'entame du processus ; les utilisations d'une évaluation peuvent aller au-delà de son objectif, puisque d'autres parties prenantes ont accès aux données de l'évaluation et les utilisent à leurs propres fins.

Dans l'espace Métadonnées, veuillez fournir toute information supplémentaire qui contribuera à compléter et comprendre les données requises dans cette partie du questionnaire.

SECTION 8: Dissémination des données



8.1. Niveau de détail auquel les données sont rapportées

Afin de rendre les données pertinentes pour l'élaboration, le suivi et l'évaluation des politiques, il faut procéder à une analyse en fonction de différents niveaux d'unité administrative et groupes de population. Par exemple, il est nécessaire de réaliser une analyse selon le sexe afin de contrôler la parité entre les apprenants masculins et féminins. Veuillez indiquer tout ce qui est applicable à cette évaluation et fournir des informations complémentaires dans les Métadonnées à la fin de la partie.



8.2. Plateformes de dissémination

De nombreuses plateformes peuvent servir à informer le public et la communauté éducative sur les résultats et la disponibilité des données d'une évaluation. Veuillez sélectionner OUI pour toutes les plateformes utilisées afin de diffuser les données de l'évaluation concernée.



8.3. Accessibilité aux données

Les microdonnées ne sont pas facilement accessibles, notamment dans le cadre des évaluations des apprenants. Cette sous-section vise à collecter toutes les informations pratiques qui devraient être connues des chercheurs ou toute personne intéressée à l'utilisation des microdonnées et des documents afférents pour mener des recherches. Cette information n'est pas souvent documentée. Veuillez-vous appuyer sur les responsables indiqués afin de collecter cette information ou valider les informations qui pourraient être disponibles pour vous assurer qu'elles sont toujours d'application, et complétez les cellules appropriées.



8.4. Défis

L'administration et la gestion des évaluations à grande échelle représentent toujours un défi, notamment dans les pays en développement. Ces défis sont généralement rapportés pour une prise en compte ultérieure. Veuillez fournir un bref résumé de ces défis dans l'espace approprié, en rapport avec les phases suivantes du processus d'évaluation des acquis scolaires: planification, développement des épreuves, collecte des données, traitement des données, analyse des données, réforme des politiques en éducation, dissémination des données et autres.



Veuillez-vous reporter au glossaire pour la définition de **diffusion des données**.

5. GLOSSAIRE ET DÉFINITIONS

Année d'études : Un cycle spécifique d'instruction au niveau de l'enseignement initial correspondant généralement à une année scolaire. Les apprenants de la même année d'études ont généralement un âge similaire. Les termes « classe » ou « cohorte » sont aussi utilisés (glossaire ISU).

Trait/construit : une caractéristique hypothétique non observable ou mentale qui est utilisée afin d'expliquer les résultats d'un individu à une évaluation. Il est mesuré uniquement au travers de résultats d'observations ou de tâches desquelles on déduit le niveau de la personne évaluée. Les construits ne peuvent pas être observés ou mesurés directement.

CITE: Système de classification fournissant un cadre pour la description statistique complète des systèmes éducatifs nationaux et une méthodologie permettant de traduire les programmes éducatifs nationaux en niveaux d'éducation comparables sur le plan international. Le programme d'enseignement est l'unité de classement de base de la CITE. Les programmes y sont également classés selon le domaine d'études, l'orientation et les débouchés des programmes. Veuillez cliquer sur le lien ci-dessous pour plus d'information sur la classification CITE.

<http://www.ISU.unesco.org/Enseignement/Pages/international-standard-classification-of-enseignement.aspx>

Codage des données : le processus qui consiste en général à assigner des valeurs numériques aux réponses exprimées au départ sous forme numérique, textuel, sonore ou visuelle. Toutefois, les logiciels d'analyse qualitative utilisent aussi des codes sous forme de texte pour faire des analyses quantitatives. Le principal objectif est de faciliter le traitement automatique des données dans un but analytique.

Compétence : La capacité à mobiliser et à exploiter des ressources internes telles que les connaissances, les aptitudes et les comportements, ainsi que des ressources externes telles que des bases de données, des collègues, des pairs, des bibliothèques, des instruments etc., en vue de résoudre efficacement des problèmes spécifiques dans des situations réelles.

Diffusion des données : la distribution des données à l'intention des utilisateurs au travers de divers médias (nouveaux et traditionnels) tels que des médias internet ou en ligne, des conférences de presse ou des publications dans la presse, des articles dans des journaux imprimés, des interviews à la télévision ou à la radio, etc.

Échantillonnage non probabiliste :

Échantillonnage au jugé : on prélève un échantillon en se fondant sur certains jugements au sujet de l'ensemble de la population. L'hypothèse qui sous-tend son utilisation est que l'enquêteur sélectionnera des unités qui représentent la population.

Échantillonnage de commodité ou à l'aveuglette : on ne sélectionne des unités d'échantillonnage que si on peut y avoir facilement et commodément accès.

Échantillonnage par quotas : il s'effectue jusqu'à ce qu'un nombre précis d'unités (de quotas) pour diverses sous-populations ait été sélectionné. Puisqu'il n'existe aucune règle qui régirait la façon dont il faudrait s'y prendre pour remplir ces quotas, l'échantillonnage par quotas est réellement un moyen de satisfaire aux objectifs en matière de taille d'échantillon pour certaines sous-populations.

Échantillonnage volontaire : ce type d'échantillonnage intervient lorsque des gens offrent volontairement leurs services pour l'étude dont il est question.

Échantillonnage probabiliste :

Échantillonnage aléatoire simple : chaque membre d'une population a une chance égale d'être inclus dans l'échantillon. Il est indispensable de dresser une liste de toutes les unités de la population observée pour sélectionner un échantillon aléatoire simple.

Échantillonnage à plusieurs degrés : ressemble à la méthode d'échantillonnage en grappes, sauf qu'il faut dans ce cas prélever un échantillon à l'intérieur de chaque grappe sélectionnée, plutôt que d'inclure toutes les unités dans la grappe. Ce type d'échantillonnage exige au moins deux degrés. On identifie et sélectionne au premier degré de grands groupes ou de grandes grappes. Ces grappes renferment plus d'unités de la population qu'il n'en faut pour l'échantillon final. Pour obtenir un échantillon final, on prélève au second degré des unités de la population à partir des grappes sélectionnées (à l'aide de l'une des méthodes d'échantillonnage probabiliste possibles). Si l'on utilise plus de deux degrés, le processus de sélection d'unités de la population à l'intérieur des grappes se poursuit jusqu'à l'obtention d'un échantillon final.

Échantillonnage à plusieurs phases :

Un échantillonnage à plusieurs phases entraîne la collecte de données de base auprès d'un échantillon d'unités de grande taille et ensuite, pour un sous-échantillon de ces unités, la collecte de données plus détaillées. La forme la plus courante d'échantillonnage à plusieurs phases est l'échantillonnage à deux phases (ou l'échantillonnage double), mais il est également possible d'effectuer un échantillonnage à trois phases ou plus.

L'échantillonnage à plusieurs phases est assez différent de l'échantillonnage à plusieurs degrés, malgré les similarités entre eux sur le plan de leur appellation. Même si l'échantillonnage à plusieurs phases suppose aussi le prélèvement de deux échantillons ou plus, dans son cas, tous les échantillons sont tirés de la même base de sondage et les unités sont structurellement les mêmes à chaque phase. Comme dans le cas de l'échantillonnage à plusieurs degrés, plus l'on utilisera de phases, plus le plan d'échantillonnage et l'estimation deviendront complexes.

Échantillonnage avec probabilité proportionnelle à la taille : pour l'échantillonnage probabiliste, il faut que chaque membre de la population observée ait une chance d'être inclus dans l'échantillon, mais il n'est pas nécessaire que cette probabilité soit la même pour tous. Si la base de sondage contient des informations sur la taille de chaque unité (comme le nombre d'apprenants de chacun des établissements scolaires ou classe) et si la taille de ces unités varie, on peut utiliser cette information dans le cadre de la sélection de l'échantillonnage afin d'en accroître l'efficacité. Cela s'appelle l'échantillonnage avec probabilité proportionnelle à la taille (PPT). Dans le cas de cette méthode, plus la taille de l'unité est grande, plus sa chance d'être incluse dans l'échantillon est élevée. Il faut que la mesure de la taille soit exacte pour que cette méthode accroisse l'efficacité.

Échantillonnage en grappes : la technique de l'échantillonnage en grappes entraîne la division de la population en groupes ou en grappes comme son nom l'indique. Suivant cette technique, on sélectionne au hasard un certain nombre de grappes pour représenter la population totale, puis on englobe dans l'échantillon toutes les unités incluses à l'intérieur des grappes sélectionnées. On n'inclut dans l'échantillon aucune unité de grappes non sélectionnées; ces unités sont représentées par celles tirées de grappes sélectionnées. La technique en question diffère de la technique d'échantillonnage stratifié, qui entraîne la sélection d'unités de chaque groupe.

Échantillonnage stratifié : lorsqu'on utilise l'échantillonnage stratifié, on divise la population en groupes homogènes (appelés strates), qui sont mutuellement exclusifs, puis on sélectionne à partir de chaque strate des échantillons indépendants. On peut utiliser n'importe quelle méthode d'échantillonnage pour sélectionner l'échantillon à l'intérieur de chaque strate. La méthode d'échantillonnage peut varier d'une strate à une autre. Lorsqu'on utilise l'échantillonnage aléatoire simple pour sélectionner l'échantillon à l'intérieur de chaque strate, on appelle le plan d'échantillonnage un plan d'échantillonnage aléatoire simple stratifié. On peut stratifier avant l'échantillonnage une population au moyen de toute variable dont on dispose pour la totalité des unités incluses dans la base de sondage (comme l'année d'études, le sexe, l'âge, la province/région/zone de résidence, le propriétaire de l'établissement scolaire).

Échantillonnage systématique ou par intervalle : signifie qu'il existe un écart, ou un intervalle, entre chaque unité sélectionnée qui est incluse dans l'échantillon. Vous devez suivre les étapes énumérées ci-dessous pour sélectionner un échantillon systématique:

- i. Numéroté de 1 à N les unités incluses dans votre base de sondage (où N est la taille de la population totale).
- ii. Déterminer l'intervalle d'échantillonnage (K) en divisant le nombre d'unités incluses dans la population par la taille de l'échantillon que vous désirez obtenir.
- iii. Sélectionner au hasard un nombre entre 1 et K. Ce nombre s'appelle l'origine choisie au hasard et serait le premier nombre inclus dans votre échantillon.
- iv. Sélectionner chaque K^e (dans ce cas, chaque 4^e) unité après ce premier nombre.

Édition des données : des contrôles visant à identifier des données incohérentes, inexacts et manquantes en vue de leur correction ou de leur remplacement grâce à un suivi avec le répondant, une révision manuelle des données ou une imputation.

Enjeu de l'évaluation

Évaluation à enjeu élevé : une évaluation à enjeu élevé est une évaluation qui a des conséquences importantes pour la **personne évaluée**, sur la base de ses résultats. La réussite apporte des avantages importants, tels que passer dans l'année d'études supérieure, obtenir un diplôme d'études secondaires ou une bourse d'études, accéder au marché du travail ou obtenir un permis pour exercer une profession. L'échec a lui aussi des conséquences, telles que l'obligation de suivre des cours de rattrapage ou l'impossibilité d'exercer une profession. Parmi les exemples de tests aux enjeux élevés, on trouve les examens d'entrée à l'université, les examens de fin d'études secondaires et les examens d'admission à une profession.

Évaluation à faible enjeu : elle est destinée à contrôler si le système d'éducation (ou le système scolaire) fonctionne bien. Les évaluations à faible enjeu n'ont pas de conséquences directes sur les personnes évaluées. Par exemple, quels que soient leurs scores, les apprenants ne seront pas retenus dans une année d'études, on ne leur refusera pas l'accès à un niveau d'études supérieur ou ils ne seront pas orientés vers des programmes plus ou moins académiques. Néanmoins, toutes les parties prenantes peuvent ne pas bénéficier cette neutralité. Les enjeux peuvent être élevés pour les enseignants et les établissements scolaires si ceux dont les élèves obtiennent des résultats médiocres sont pris pour cible par les autorités compétentes et doivent subir diverses actions (licenciement, formation complémentaire, etc.) Les agences gouvernementales peuvent se trouver prises sous les feux des partis d'opposition lorsque les résultats sont faibles ou en baisse. Dès lors, si les enjeux de l'évaluation sont faibles pour les apprenants, ils peuvent s'avérer élevés pour d'autres.

Enquête pilote : une enquête menée auprès de quelques individus de la population cible ou de l'échantillon d'une enquête, en vue de tester et affiner les outils de l'enquête (questionnaire et manuel, manuel de traitement des données et programmes) avant de réaliser la principale collecte de données auprès de l'ensemble de la population cible ou de l'échantillon complet.

Évaluation des résultats de l'apprentissage : se réfère à une mesure de la réussite dans l'atteinte d'objectifs d'apprentissage. Elle peut se concentrer sur des programmes scolaires spécifiques, exploiter diverses méthodes d'évaluation (tests/examens écrits, oraux ou pratiques, projets et dossiers de présentation) et être menée pendant ou au terme d'un programme d'enseignement.

Évaluation obligatoire : une **évaluation aux enjeux élevés** menée parmi tous les apprenants d'une année d'études donnée. Dans le cadre de cette enquête, voici des exemples d'évaluations obligatoires : le General Certificate of Education Ordinary level (GCE-OL) ou Advanced level (GCE-AL) dans le système britannique, le Baccalauréat (Bac) dans le système français, le Diplôme d'étude collégiale (DEC) au Québec-Canada. Seuls les apprenants titulaires du GCE-AL, du Bac, du DEC, brevet ou autres diplômes équivalents sont autorisés à accéder à l'enseignement supérieur ou à passer des tests en vue d'une sélection pour accéder à l'enseignement secondaire de second cycle.

Évaluation organisée à l'échelle de l'école : une évaluation des apprenants organisée et menée régulièrement par chaque établissement scolaire d'un pays. Les outils d'évaluation sont généralement mis au point par le personnel enseignant de l'établissement. Les résultats sont utilisés en vue de fournir un retour d'informations direct aux apprenants et aux parents, de connaître le niveau de la classe et d'améliorer le processus d'enseignement-apprentissage. Dans certains pays, on prend en

compte les scores obtenus à ces évaluations (pondération par rapport au score final total) pour l'obtention d'un diplôme ou la sélection des apprenants.

Format du stimulus- Texte (source : OECD 2015 PISA reading framework¹, Pages 17-18)

Textes continus : Les textes continus ou suivis sont formés de phrases organisées en paragraphes. Parmi les exemples de textes suivis, on trouve des articles de journaux, reportages, romans, nouvelles, rapports et lettres.

Textes mixtes : Les textes mixtes sont des objets cohérents consistant en un ensemble d'éléments sous format continu et non continu. Dans des textes mixtes bien construits, les composants (par exemple, une explication en prose accompagnée d'un graphique ou d'un tableau) se renforcent mutuellement au travers de liens de cohérence et de cohésion au niveau local et global. On trouve communément des textes mixtes dans des magazines, des ouvrages et des rapports de référence, les auteurs faisant appel à divers formats afin de communiquer l'information.

Textes multiples: Les textes multiples sont des textes qui ont été générés de manière autonome et qui ont du sens de manière autonome ; ils sont juxtaposés pour une occasion spécifique ou ont été associés dans le cadre de l'évaluation. La relation entre les textes peut ne pas être évidente ; ils peuvent être complémentaires ou à sens opposé. Par exemple, un ensemble de sites web de différentes agences de tourisme peuvent donner des conseils similaires ou non aux touristes. Les textes multiples peuvent se caractériser par un seul format « pur » (par exemple, continu), ou combiner textes continus et textes non continus.

Textes non continus : Les textes non continus, également dénommés documents, sont organisés différemment des textes continus, de sorte qu'ils permettent au lecteur d'employer des stratégies différentes pour comprendre l'information écrite. Quelques exemples de textes non continus : listes, tableaux, graphiques, diagrammes, annonces, horaires, catalogues et formulaires. Ces textes peuvent être fixes ou dynamiques.

Besoins spéciaux : ce terme couvre les déficiences, les limitations des activités et la participation. Il s'agit d'un phénomène complexe qui découle de l'interaction entre les caractéristiques corporelles d'une personne et les caractéristiques de la société où elle vit. Pour surmonter les difficultés auxquelles les personnes handicapées sont confrontées, des interventions destinées à lever les obstacles environnementaux et sociaux sont nécessaires (Organisation mondiale de la santé²).

Autisme : trouble envahissant du développement humain affectant significativement la communication verbale et non verbale et l'interaction sociale, généralement manifeste avant l'âge de trois ans, ayant un impact négatif sur les performances d'apprentissage d'un enfant. Il se caractérise également par des activités répétitives et des mouvements stéréotypés, une résistance au changement de l'environnement ou de routines quotidiennes et des réactions inhabituelles à des expériences sensorielles. Le terme autisme ne s'applique pas si les performances d'apprentissage de l'enfant sont principalement affectées par des troubles émotionnels. Si un enfant présente les caractéristiques de l'autisme après l'âge de trois ans, il peut être diagnostiqué comme étant autistique si les critères ci-dessus sont remplis (Congrès américain, IDEA 1997).

Déficience auditive (incluant la surdité) : trouble de l'audition, permanent ou non, affectant les performances d'apprentissage d'un enfant mais qui n'est pas inclus dans la définition de la « surdité » (Congrès américain, IDEA 1997). **Surdi-cécité** : désigne une déficience concomitante [simultanée] de l'audition et de la vision, dont la combinaison entraîne des problèmes tellement sévères en matière de communication, de développement et d'éducation qu'il est impossible d'accueillir ces enfants dans des programmes d'enseignement spécial ouverts aux enfants uniquement atteints de surdité ou de cécité (Congrès américain, IDEA 1997).

Déficience visuelle : troubles de la vision qui, même s'ils sont corrigés, affectent les performances d'apprentissage d'un enfant. L'expression inclut la malvoyance et la cécité (Congrès américain, IDEA 1997).

¹ <http://www.oecd.org/pisa/pisaproducts/Draft%20PISA%202015%20Reading%20Framework%20.pdf>

² <http://www.who.int/topics/disabilities/fr/>

Effet de l'environnement: cela se produit lorsque le candidat à l'examen/évaluation est affecté par les caractéristiques de l'environnement dans lequel le test est administré. Par exemple, il peut être appelé à prendre un autre test ou à le passer à un autre moment.

Surdité : désigne une déficience auditive tellement grave qu'un enfant est handicapé dans le traitement des informations linguistiques liées à l'audition, avec ou sans amplification, ce qui a un impact négatif sur ses performances d'apprentissage (Congrès américain, IDEA 1997).

Troubles émotionnels : un état présentant une ou plusieurs des caractéristiques suivantes sur une longue durée et dans une mesure suffisante pour affecter les performances d'apprentissage d'un enfant : incapacité d'apprendre qui ne peut pas s'expliquer par des facteurs intellectuels, sensoriels ou de santé - incapacité à développer ou entretenir des relations interpersonnelles satisfaisantes avec des pairs et des enseignants – comportements ou sentiments inappropriés dans des circonstances normales – sentiment général persistant de tristesse ou de dépression – tendance à développer des symptômes physiques ou des peurs associées à des problèmes personnels ou scolaires - Ce terme inclut la schizophrénie. Elle ne s'applique pas aux enfants vivant des situations d'inadaptation sociale, sauf s'il a été déterminé qu'ils souffraient de troubles émotionnels (Congrès américain, IDEA³ 1997).

Trouble de la parole : trouble de la communication tel que le bégaiement, des troubles articulatoires ou vocaux affectant les performances d'apprentissage d'un enfant (Congrès américain, IDEA 1997).

Imputation : une procédure statistique utilisée afin de déterminer les valeurs de remplacement pour des données manquantes, inexactes ou incohérentes. Quelques exemples de méthodes d'imputation fréquemment utilisées : « hot deck », « cold deck », exclusion par liste ou par paire, imputation moyenne, imputation par régression, projection de la dernière observation, imputation par régression stochastique et imputations multiples.

Item : une question ou tâche unique dans une évaluation.

Note : Un item présente trois parties : le stimulus qui fournit le contexte de la tâche en format audio, vidéo ou texte, une description claire de la tâche à réaliser sur la base des stimuli, et les réponses possibles dans un format donné (dichotomiques, choix multiples, hiérarchie partielle, ouvert, rédaction, etc.)

Modèle de mesure :

Théorie classique des tests (TCT): théorie de mesure consistant en un ensemble de postulats sur les relations entre scores réels ou observés à un test et les facteurs affectant les scores. Elle est utilisée afin de mesurer et de gérer les données de performance au test et aux items. Contrairement à la théorie de la réponse d'item, elle comprend un ensemble de méthodes psychométriques plus traditionnelles (DFID⁴). Dans la TCT, le niveau global de compétence d'une personne évaluée dépend de la difficulté générale du test et vice-versa. Il est impossible d'obtenir les paramètres de chaque item d'un test ou le niveau de compétence de la personne évaluée dans chaque item sur une échelle continue de performance comme dans les modèles TRI ou de Rasch.

Modèle de Rasch : groupe de modèles mathématiques permettant de mettre en rapport et de prévoir le score d'un individu pour un item donné avec son niveau de performance sur une échelle relative à la compétence ou la caractéristique mesurée, et le paramètre de difficulté de l'item. Le modèle de Rasch considère que la probabilité qu'un individu fournisse une réponse exacte ou réalise efficacement une tâche donnée dépend uniquement de la différence entre son niveau de compétence et le niveau de difficulté de l'item. Cette probabilité augmente lorsque la compétence est supérieure à la difficulté de l'item, mais s'élève à 50 % si ces deux paramètres sont égaux.

³ **IDEA:** Individuals with Disabilities Education Act (US Congress, 1997)

⁴ <http://www.heart-resources.org/wp-content/uploads/2012/04/04-Learning-Outcomes-How-To-Note.pdf>

Théorie de la réponse d'item (TRI): Un groupe de modèles mathématiques permettant de mettre en rapport et de prévoir le score d'un individu pour un item donné avec son niveau de performance sur une échelle relative à la compétence ou à la caractéristique mesurée, et les paramètres caractéristiques de l'item : supposition, discrimination et difficulté.

Note: comparés à la théorie classique des tests (TCT), les deux modèles TRI et de Rasch ont le potentiel de donner des probabilités concernant le niveau de compétence de la personne évaluée indépendamment des items, ainsi que des probabilités concernant des caractéristiques de l'item également indépendamment du groupe de test. C'est ce que l'on appelle le principe d'invariance.

Méthode d'ajustement de la non réponse : méthode statistique visant à remplacer des données manquantes ou inexactes. Pondération et imputation constituent des exemples de méthode d'ajustement à la non réponse généralement utilisées pour tenir compte des non réponses avant de réaliser des analyses.

Méthode de référencement de la performance

Référencement critériel (Critèrien-referenced assessment, CRA) : lorsque les apprenants sont évalués en fonction d'une norme ou de résultats prédéfinis, de résultats ciblés ou souhaitables, d'une référence ou d'un critère, l'évaluation est dite critériel. Dans les CRA, tous les apprenants sont évalués par rapport au critère ou à la référence, qui peut être un ensemble spécifique de connaissances ou de compétences. Dans l'enseignement, les CRA sont généralement conçues de manière à déterminer si l'apprenant maîtrise la matière enseignée dans une année d'études ou un cours spécifiques.

Évaluation individuelle/ipsative : elle consiste à évaluer les résultats d'une personne à un instant donné par rapport à ses résultats passés, en vue de déterminer si des progrès ont été réalisés. Cette pratique nécessite des données longitudinales.

Évaluation normative (Norm-referenced assessment, NRA): Elle consiste à comparer une personne évaluée à ses pairs, ce qui signifie que les scores des apprenants sont simplement classés par ordre de grandeur et le classement de chaque apprenant est comparé à celui des autres. On ne tente pas d'interpréter les scores et de savoir ce que les apprenants connaissent ou peuvent faire, excepté dans le sens limité où les performances d'un apprenant sont semblables à celles d'apprenants ayant des performances faibles, moyennes ou élevées dans le groupe.

Déclaration typique de résultat individuel : *les apprenants ont obtenu un score d'échelle de 450 (sur une moyenne d'échelle de 500, écart type 100).*

Déclaration typique de résultat de groupe : *le score moyen sur l'échelle de maîtrise pour l'année d'études X dans l'évaluation Y est S et l'écart type D.*

Notes: 1-Dans les tests normatifs, les scores sont souvent exprimés en percentiles : un apprenant dont le score est X reçoit le percentile P où P% des apprenants ont des scores de X ou moins. Il est donc clair que les tests normatifs sont conçus en vue de trier et classer les apprenants « sur la courbe » et non de vérifier s'ils satisfont à une norme ou un critère. Dès lors, il ne faut pas recourir aux évaluations normatives en vue d'évaluer si des apprenants ont satisfait à des normes souhaitables (ou minimales). 2- En théorie, un test peut être à la fois critériel et normatif. Une CRA peut être utilisée afin de mesurer l'apprentissage des apprenants par rapport à des normes, avec des scores de références spécifiques sur l'échelle choisie afin de distinguer les niveaux de satisfaction aux normes. Ensuite, il est possible de classer les scores. Toutefois, le degré de précision de la CRA peut s'avérer insuffisant pour réaliser les objectifs de la NRA. On peut évaluer une NRA pour l'utiliser en tant que CRA, mais son contenu et sa précision pourraient être insuffisants en vue de déterminer les niveaux de connaissances et de compétences. Un test ne devrait viser les deux objectifs que si ceux-ci sont clairement définis au cours du développement du test.

Métrique Réussite/Échec : Les décisions réussite/échec sont déterminées par des scores sur l'échelle de compétence ainsi que des scores exprimés en pourcentage. Le format de présentation réussite/échec est une variante de la présentation du niveau de compétences qui ne compte que deux niveaux : réussite ou échec. Parfois, on

subdivise le groupe « réussite » en sous-groupe « réussite avec distinction », « réussite première division », etc. **Dans le cadre du catalogue, toutes ces subdivisions sont regroupées dans la catégorie « réussite ».**

Déclaration typique de résultat individuel : L'apprenant a réussi le test. L'apprenant a échoué au test
Déclaration typique de résultat de groupe : 22 % des apprenants ont réussi le test.

Note : Habituellement, la décision relative à la réussite ou à l'échec se fonde sur un score pivot ou un score en pourcentage pour le test. Par exemple, les scores pivots sont prédéfinis comme étant la performance minimale acceptable, soit 50 % ou 80 %. Il n'existe pas de règle générale quant au pourcentage pivot, étant donné qu'il dépend de la difficulté et du contenu des items du test et de l'objectif de la décision réussite/échec. Sur le plan opérationnel, la réussite peut également être fonction d'un nombre fixe de places disponibles. Par exemple, s'il y a seulement 30 places disponibles et 1 000 personnes évaluées, on considère que les 3 % ayant les meilleurs résultats auront réussi l'examen, quels que soient leurs scores, si les 30 places disponibles doivent être pourvues. On peut imaginer une combinaison : un score de performance minimal et un nombre maximum de places disponibles ; cela signifie que parmi ceux qui obtiennent un score minimal Y, un maximum de 30 réussiront ; si moins de 30 personnes évaluées atteignent le score minimal, le responsable de l'évaluation peut réduire le score minimal ou ne pas pourvoir toutes les places disponibles (par exemple, dans des établissements d'enseignement secondaire de deuxième cycle).

Métrique en pourcentage : Le nombre d'items du test auxquels un apprenant répond correctement, divisé par le nombre total d'items du test et multiplié par 100 égale le score en pourcentage de cet apprenant. Le score est donc un pourcentage et non le nombre de réponses correctes. Si un test présente des items notés différemment, le score en pourcentage de l'apprenant est le total des points sur tous les items, divisé par le total des points possibles, multiplié par 100.

Déclarations typiques de résultat individuel : les apprenants ont répondu à X % des items du test correctement. Les apprenants ont obtenu X % des points possibles au test.
Déclarations typiques de résultat de groupe : le pourcentage moyen des réponses correctes du groupe s'élevait à X % avec un écart type de Y %. Le pourcentage des points obtenus allait de X % à Y % avec une moyenne de Z %.

Scores d'échelle : Les scores de l'échelle sont des estimés des habiletés générés par les modèles de la théorie de la réponse à l'item (TRI), sur base des vecteurs de réponse des apprenants. Les scores d'échelle sont destinés à procurer une métrique cohérente pour différentes versions d'un test et dans le temps.

Métrique de niveau de compétences : il s'agit de la classification des apprenants par fourchettes de performance identifiées par séries de scores pivots sur l'échelle de performance. Les niveaux de compétence sont souvent utilisés dans les tests selon des critères. Pour chacun de ces niveaux, il faut une description de la signification du niveau en question, en termes de connaissances, compétences, attitudes, etc. Chaque niveau (ou fourchette) représente un degré de maîtrise de l'objectif de mesure du test. Les mêmes niveaux de maîtrise sont désignés par des mots ou des lettres, par exemple, « inférieur au basique », « basique », « supérieur au basique » ; ou « élevé », « moyen », « bas » ; ou A, B, C.

Déclarations typiques de résultat individuel : l'apprenant se classe au niveau 2 ou basique. L'apprenant se classe au niveau 4 ou avancé.
Déclarations typiques de résultat de groupe : dans les établissements scolaires de ce type, X % des apprenants appartiennent au niveau 1 (inférieur au niveau de base), Y % au niveau 2 (niveau de base), Z % au niveau 3 (Compétent) et W % au niveau 4 (Avancé). Dans cette école, X % des apprenants se trouvent au niveau 2 (Compétent) ou supérieur.

Note : on associe généralement niveaux de compétences et scores sur l'échelle des compétences. La détermination des scores pivots d'une échelle de test spécifique commence par la définition et la description de ce que l'on entend par les niveaux de compétences en termes de connaissances et de compétences, évalue les items de l'échelle en termes de contenu et de complexité, et poursuit par des jugements pour déterminer systématiquement où placer les scores pivots.

Métrique de présentation des résultats : les différentes formes de présentation des résultats de l'évaluation des résultats d'apprentissage, ou des acquis scolaires, à des fins de communication aux parties prenantes du monde de l'éducation, y

compris les apprenants. Les résultats peuvent être présentés par individus ou agrégés pour des groupes spécifiques. Les formes possibles (métrique) de présentation des résultats sont le pourcentage de réussite ou d'échec, l'échelle et les niveaux de compétences.

Non réponse du candidat : statut des répondants qui est déterminé lorsque les participants sélectionnés n'ont pas complété le questionnaire ou lorsque le taux d'achèvement du questionnaire n'est pas suffisant pour prendre en compte ses données à des fins d'analyse.

Plan d'échantillonnage

Échantillon : sous-ensemble d'individus d'une population spécifique, formé conformément à un processus de sélection (avec ou sans remplacement et de façon aléatoire ou non). Un échantillon formé par une sélection aléatoire d'individus sur la base de probabilités connues est qualifié d'échantillon aléatoire ou probabiliste. Un échantillon non aléatoire est formé sur la base d'une méthode de sélection subjective.

Échantillonnage : méthode de conception d'un échantillon sur la base d'un plan (d'échantillonnage) contenant tous les individus de la population cible. Les méthodes d'échantillonnage d'une population dépendent de nombreux paramètres tels que le processus de sélection de l'échantillon (avec ou sans remplacement et de façon aléatoire ou non), la structure du cadre d'échantillonnage, le niveau de ventilation des données et l'analyse requise et le budget disponible pour l'étude. Ci-dessous, des méthodes échantillonnage probabilistes et non probabilistes courantes (Statistique Canada)⁵:

Plan d'échantillonnage matriciel : le processus consistant à soumettre différents sous-ensembles d'items d'un test à différents sous-ensembles de répondants, de façon à ce que chaque item soit soumis à au moins un sous-ensemble de l'échantillon des répondants. Généralement, il y a des items importants nommés « items principaux » ou « core items en anglais » soumis à tous les répondants, et des items « fractionnés » ou « split items » en anglais qui sont soumis uniquement à un sous-ensemble de répondants. En général, les items principaux sont prédictifs de nombreux items fractionnés du test.

Recensement : enquête officielle concernant la totalité de la population d'un système défini. Par exemple, un recensement scolaire touche tous les établissements scolaires d'un système éducatif. Dans le contexte du catalogue de l'ISU, la population cible est soit les apprenants d'un âge ou d'une année d'études donnés, soit des candidats inscrits qui remplissent les conditions nécessaires pour passer un examen public donné, ou pour participer à une évaluation.

Saisie des données : le processus qui consiste à encoder les données dans un système de base de données, notamment lorsque la collecte des données n'est pas informatisée.

Score pivot : point particulier sur l'échelle des scores à un test au-dessus ou en dessous duquel les scores sont interprétés différemment.

Taux de participation : nombre total d'individus dans l'échantillon sélectionné ayant participé à l'enquête (complété le questionnaire) exprimé en pourcentage du nombre initial d'individus sélectionnés (taille de l'échantillon) ou population cible totale.

Taux de participation pondéré : un taux de participation propre à un échantillon qui rend compte de la pondération de chaque individu ayant participé à l'enquête.

⁵ <http://www.statcan.gc.ca/edu/power-pouvoir/ch13/prob/5214899-eng.htm#a1>
<http://www.statcan.gc.ca/edu/power-pouvoir/ch13/nonprob/5214898-eng.htm>

Test adaptatif informatisé (CAT) : une méthode d'administration de tests qui sélectionne les items en fonction du niveau de réussite du répondant, déterminé par les réponses précédentes. Un modèle CAT est très différent d'un test fixe informatisé car il sélectionne les items sur la base des réponses précédentes pour chaque répondant et peut enregistrer la manière dont chaque répondant recherche, conserve et récupère les informations.

Test fixe informatisé (CFT) : la version numérique du test «papier-crayon», qui vise à faciliter le traitement des données et à réduire le risque d'erreur humaine lors de la compilation d'une base de données.

Test standardisé : test dans lequel les items/tâches ou questions, les conditions d'administration, l'édition, la notation et l'interprétation des résultats s'appliquent de façon cohérente et prédéterminée pour toutes les personnes évaluées.

Traitement des données : une série d'opérations manuelles, automatiques ou électroniques telles que la validation, le tri, la réduction, l'agrégation de données. Ces opérations sont généralement suivies de l'extraction, la transformation, la classification, l'analyse et la présentation des données.

Types d'évaluation :

Évaluation internationale : il s'agit d'une évaluation des résultats de l'apprentissage qui fournit des informations similaires à celles de l'évaluation nationale, mais pour plus d'un système d'éducation nationale. Généralement, elle n'est pas sensible à des systèmes individuels puisque son principal objectif est la comparabilité des résultats entre les pays participants.

Évaluation nationale : dans le contexte du catalogue de l'ISU, il s'agit d'une évaluation des résultats de l'apprentissage qui vise à décrire les performances des apprenants à un âge ou dans une année d'études particuliers et à fournir un retour d'informations sur un nombre limité de mesures de résultat considérées comme importantes au niveau du système éducatif par les décideurs politiques et la communauté éducative au sens large. Elle est généralement menée auprès d'un échantillon d'apprenants et vise également à collecter des informations contextuelles (apprenants, enseignants et parents) importantes en vue d'associer analyse et questions politiques aux niveaux national, infranational et local.

Examens publics : cette expression désigne des examens standardisés de sortie ou de fin d'études qui sont généralement organisés par le jury d'examen d'un état central fédéralisé ou non d'un pays donné en vue de promouvoir, sélectionner ou octroyer un diplôme à **tous** les candidats qui ont réussi ou sont supposés avoir acquis et couvert, formellement ou non, le programme d'études d'un programme d'enseignement formel, faisant partie des exigences pour l'obtention d'un diplôme. L'examen public est généralement organisé chaque année, à destination de toute personne qui s'inscrit quels que soient son âge et sa profession (dans certains pays). Contrairement aux évaluations nationales, les données contextuelles des apprenants sont rarement collectées à l'occasion des examens publics.

Valeurs probables : ce sont les estimés des habiletés qui sont générés par la combinaison entre les techniques de modélisation de la théorie de la réponse à l'item et les imputations multiples à partir des modèles de régression latente. Même si les valeurs probables ne sont pas des scores réels observés, elles ont la propriété d'un score à un test qui permet de faire des analyses secondaires à partir des données collectées à l'aide des plans d'échantillonnage matriciel complexes.