
Concevoir une évaluation nationale

FICHE 3

Docimologie

Ces vingt dernières années, les évaluations nationales sont devenues un outil important de mesure de la réussite éducative. Malgré la grande variété des programmes d'évaluation nationaux, qui ont des buts et objectifs différents, ils visent pour la plupart à mesurer l'évolution des acquis scolaires sur la durée. Les évaluations nationales peuvent également être utilisées pour fournir des données destinées à suivre les progrès nationaux réalisés en vue d'atteindre [l'Objectif de développement durable 4 \(ODD 4\)](#). Un programme national d'évaluation nécessite une planification détaillée en fonction de grands éléments d'appréciation, énumérés ci-dessous.

Quel est le but visé par l'évaluation ?

D'après l'Institut de statistique de l'UNESCO (ISU), la plupart des évaluations nationales ont deux buts généraux :

1. « contrôler dans quelle mesure les élèves atteignent les objectifs d'apprentissage clés figurant dans le programme national et favoriser l'apprentissage pour tous » et
2. « demander des comptes aux établissements scolaires et donner aux élèves et à leurs parents des informations sur leurs progrès d'apprentissage » (ISU, 2017, p. 7).

Quelles compétences seront testées ?

Les évaluations peuvent être conçues pour tester des compétences générales couvrant diverses disciplines (comme les compétences de lecture, écriture, calcul, résolution de problèmes ou communication), comme dans le [Programme international pour le suivi des acquis des élèves \(PISA\)](#), ou pour évaluer le programme prévu ou achevé, comme d'autres instruments. Toutes les évaluations nationales mesurent les compétences cognitives en langue/lecture-écriture et mathématiques/calcul, et, dans certains pays, elles portent aussi sur d'autres domaines comme les sciences, les études sociales et les langues. Pour évaluer ces compétences variées, l'utilisation de différents instruments, comme des exercices oraux, pratiques ou des dossiers, peut s'avérer nécessaire.

Quel que soit le domaine, il est important de développer un cadre qui définit clairement les compétences et aptitudes à tester. L'une des difficultés associées à l'évaluation de compétences transdisciplinaires est qu'il n'est pas forcément évident de se mettre d'accord sur les définitions de ces compétences. Il peut alors être utile de faire intervenir des experts externes qui aideront à les définir.

Afin de comprendre les variables qui pourraient affecter l'apprentissage, il faut prendre soin de s'interroger sur les informations générales qui doivent être recueillies et sur la manière dont elles doivent être collectées (par exemple par le biais de questionnaires destinés aux enseignants ou élèves).

Qui sont les groupes cibles de l'évaluation ?

Au moment de sélectionner un groupe cible, les pays devraient se demander si l'évaluation doit :

- Cibler **un groupe d'âge ou une classe de niveau** ? Les deux méthodes ont des avantages et des désavantages. L'un des avantages de l'échantillonnage par classe est qu'il permet d'avoir plus d'informations générales sur les pratiques pédagogiques et les conditions en classe, qui peuvent être mises en relation avec les données sur les acquis scolaires.
- Être **fondée sur un échantillon ou sur un recensement** ? Les examens et tests visant à suivre les écoles sont souvent obligatoires pour tous les élèves, tandis que les tests axés sur l'évaluation du système éducatif dans son ensemble sont souvent soumis à un échantillon représentatif.
- Inclure **les enfants non scolarisés** ? Par exemple en utilisant des enquêtes auprès des ménages.

Comment la qualité de l'instrument d'évaluation peut-elle être garantie ?

Une évaluation de la qualité se caractérise par le contenu, la validité convergente et prédictive, la fiabilité et l'équité.

La validité d'un test désigne la capacité d'un test à mesurer réellement ce qu'il vise à mesurer. La validité est généralement considérée comme l'aspect le plus important des tests éducatifs, car elle concerne le sens que l'on donne aux résultats des tests, ainsi que la confiance qu'on peut leur accorder pour mesurer les bonnes compétences. Une évaluation particulièrement valide couvre tous les aspects pertinents des performances d'un élève. Pour estimer la validité d'un test, on peut notamment utiliser les méthodes suivantes : la validation croisée, l'analyse des items, l'intercorrélation des items et l'analyse factorielle.

La fiabilité d'un test est la capacité d'une évaluation à produire des résultats stables et cohérents. Un niveau de fiabilité adéquat est une condition nécessaire de la validité d'un test : si la mesure n'est pas fiable, elle ne peut pas être valide. La théorie de réponse aux items (IRT), qui applique de nouvelles méthodes de mise à l'échelle, envisage la fiabilité des tests sous un nouvel angle, en prenant en compte le fait que chaque item peut avoir un niveau de difficulté différent. Avec les méthodes IRT, la fiabilité du test correspond approximativement à la précision de la mesure à différents niveaux de la compétence mesurée. L'inverse de la fiabilité est l'erreur de mesure. L'exactitude et la précision de la mesure sont donc d'une importance primordiale pour garantir la plus grande fiabilité du test dans son ensemble.

L'équité d'une évaluation signifie qu'elle est exempte de toute forme de partialité. Chaque test doit convenir à toutes les personnes testées, quels que soient leur origine ethnique, leur religion, leur genre ou leur âge. Une évaluation ne doit pas désavantager une personne pour toute autre raison que l'insuffisance des connaissances et compétences que l'évaluation vise à mesurer. Pour qu'un test réponde aux exigences de validité, de fiabilité et d'équité, ses items doivent être pilotés et

analysés à l'aide de méthodes psychométriques avant d'être utilisés.

Quel doit être le format des tests ?

Pour qu'il soit valide, un test doit être composé d'items représentant l'ensemble du domaine testé. Il doit aussi contenir suffisamment d'items pour chaque niveau d'aptitude. Les items peuvent être des questions à choix multiples ou ouvertes, ou une combinaison des deux. Les questions ouvertes nécessitent un manuel de notation très strict, ainsi qu'une formation approfondie des évaluateurs. De nombreux pays abandonnent les tests sur papier et se tournent vers les tests sur ordinateur. Cela permet d'envisager des tests adaptatifs, c'est-à-dire des tests qui s'adaptent automatiquement au niveau d'aptitude de l'élève, favorisant une mesure plus précise de l'ensemble de la compétence, ainsi que des tests ciblés.

Un système de tests tournants (échantillonnage matriciel) est souvent utilisé dans le cas des tests fondés sur des échantillons pour suivre l'ensemble d'un système éducatif. Dans ce système tournant, le test se compose de blocs, souvent dans une série de livrets, chaque bloc représentant une partie seulement de l'ensemble du test. Chaque élève ne répond qu'à un livret, qui peut contenir n'importe quel bloc distinct de documents. Cela permet de tester une grande série d'items sans que l'évaluation soit trop longue pour les élèves. Cependant, cette méthode ne permet pas de fournir des résultats individuels pour chaque élève.

Comment mesurer précisément les tendances sur la durée ?

Pour suivre les tendances d'acquisition des apprentissages sur la durée, le test doit contenir une série d'items d'ancrage, qui se répètent à chaque cycle. Les items d'ancrage sont un moyen de garantir que les niveaux d'aptitude annoncés représentent le même niveau de difficulté au fil du temps – en d'autres termes, que les résultats numériques correspondent toujours aux mêmes niveaux de compétences. Les items d'ancrages doivent être gardés secrets pour garantir les mêmes conditions d'évaluation sur la durée.

Qui doit mettre en œuvre le test et à quelle fréquence ?

Il est important de prendre en compte la manière dont l'évaluation sera mise en œuvre. Les pays peuvent se demander s'il est préférable que l'évaluation soit menée par un ministère du gouvernement ou un groupe indépendant de spécialistes, et s'il faut confier la passation du test à un administrateur externe qualifié ou des enseignants. L'objectif de l'évaluation doit également déterminer la fréquence à laquelle elle est organisée et le moment de l'année scolaire auquel elle est prévue, par exemple en début ou en fin d'année scolaire.

Comment et à qui les résultats doivent-ils être présentés ?

Les évaluations doivent répondre à des questions importantes d'orientation politique et leurs résultats doivent être présentés à la fois aux décideurs et au grand public. Toutefois, le partage public de données détaillées jusqu'au niveau de l'établissement peut être contestable. Si des rapports sur les données des établissements sont rendus publics, il faut veiller à ce que les élèves ne puissent pas être identifiés individuellement. Il peut être nécessaire de préparer plusieurs rapports – certains plus détaillés que d'autres – en vue de présenter leurs conclusions à différents publics comme les responsables politiques, les enseignants et le grand public. Les résultats des évaluations nationales sont en général publiés dans des rapports nationaux. En fonction de la conception et de

l'objectif du test, les résultats peuvent soit être présentés sous la forme d'un score total, soit ventilés en sous-groupes qui représentent différents sous-domaines et niveaux d'aptitude.

L'expertise nécessaire est-elle disponible ?

Le développement de tests nationaux nécessite une expertise relative à la fois aux programmes et à des contenus spécifiques, ainsi que des compétences psychométriques. Certains pays disposent d'instituts nationaux ou de centres d'évaluation qui apportent l'expertise nécessaire. Il existe aussi des instituts d'évaluation nationaux et internationaux qui peuvent aider les pays et renforcer leurs capacités.

Outils

- UNESCO-ISU. 2017. [Quick guide no. 3: Implementing a national learning assessment](#). Montréal: ISU.
- UNESCO-ISU. 2018. [Quick guide no. 2: Making the case for a learning assessment](#). Montréal: ISU.
- World Bank. [SABER: Student assessment](#).

References and sources

ACER; ACER-GEM ; UNESCO-ISU ; GAML. 2017. [Principles of good practice in learning assessment](#). Montréal: UNESCO-ISU.

Greaney, V.; Kellaghan, T. 2008. [Assessing national achievement levels in education. Vol. 1.](#) Washington, DC: Banque Mondiale.

Postlethwaite, T. N.; Kellaghan, T. 2008. [National assessments of educational achievement](#). Paris: IPE-UNESCO.

UNESCO-ISU. 2017. Quick guide no. 3: [Implementing a national learning assessment](#). Montréal: ISU.

UNESCO-ISU. 2018. Quick guide no. 2: [Making the case for a learning assessment](#). Montréal: ISU.

UNESCO-ISU. 2018. [Guide rapide des indicateurs de l'éducation pour l'ODD 4](#). Montréal : ISU.

[print](#)