

Réalisations

Séminaire sur l'analyse d'items au Maroc¹

Un séminaire de formation sur l'analyse d'items s'est tenu au Maroc à la suite de l'introduction de banques d'items pour assurer la sanction des études à l'enseignement secondaire. Des données statistiques sur les items faisant partie des banques d'items disponibles ont pu être analysées au regard de leurs qualités docimologiques. De nombreux responsables du mécanisme d'évaluation mis en place dans ce pays ont participé à cette activité de formation.

On doit tout d'abord fournir quelques renseignements généraux sur cette procédure qui consiste à sanctionner les études réalisées par les élèves. L'évaluation des apprentissages se réalise à l'aide d'épreuves constituées de questions que l'on désigne habituellement sous le terme « item ». Les résultats obtenus à la suite de l'administration d'une épreuve sont utilisés pour prendre des décisions variées. Il peut s'agir de promouvoir un élève, de le sélectionner pour son admission dans une filière scolaire ou un emploi; les résultats peuvent aussi permettre de planifier des programmes d'intervention scolaire. Ces résultats peuvent aussi être utilisés dans le cadre d'un processus de « sanction » des études accomplies ou bien pour la « certification » d'une formation. On pourrait énumérer plusieurs autres utilisations des résultats obtenus à la suite d'une évaluation des apprentissages. Aussi, faut-il que les instruments utilisés pour recueillir les renseignements qui serviront à la réalisation de l'évaluation respectent de hauts standards de qualité.

Il existe des techniques d'analyses statistiques permettant d'apprécier les qualités docimologiques d'une épreuve. Nous allons faire la présentation de quelques-unes. De plus, nous rappellerons certains concepts liés aux techniques présentées. Il sera question de l'analyse d'items par rapport à leurs

¹ Plusieurs séminaires semblables ont été dispensés dans d'autres pays africains.

qualités métriques ainsi que par l'absence de biais pénalisant certains sous-groupes d'élèves touchés par l'évaluation.

Ce séminaire de formation a permis de présenter les techniques habituellement retenues pour mieux se prononcer sur les qualités métriques des épreuves par rapport aux deux techniques suivantes : i) l'analyse d'items; ii) et l'analyse du fonctionnement différencié des items.

Il est important d'examiner les qualités d'un instrument de mesure et des items qui le composent. Il s'agit du procédé statistique que l'on désigne habituellement sous le vocable d'analyse d'items. Plusieurs techniques peuvent être utilisées pour s'assurer des qualités métriques d'un instrument de collecte des données. Deux techniques d'analyse d'items ont fait l'objet de présentation : l'analyse classique et l'analyse par la technique de réponse aux items (TRI).

L'analyse classique d'items concerne les réponses à choix multiple ou sur une échelle continue ou à crédits partiels. Elle permet l'estimation de la qualité métrique de chaque item faisant partie d'un test ou d'une épreuve. Ce type d'analyse permet l'estimation de la qualité métrique de chaque item faisant partie d'un test ou d'une épreuve. Trois séries d'indices statistiques sont ordinairement retenues; les voici sommairement présentées :

- **La difficulté** Pour chaque item faisant partie d'une épreuve expérimentée, on doit en calculer la difficulté relative; une épreuve doit contenir des questions faciles, moyennement faciles, plutôt difficiles afin de recueillir des renseignements sur l'ensemble des élèves soumis à l'évaluation. Pour les items à choix multiple, il s'agit de la proportion d'élèves ayant répondu correctement à la question. Pour les items dits à crédits partiels, il faut calculer un résultat moyen.
- **La discrimination** Les questions posées aux élèves évalués doivent permettre de bien identifier ceux qui maîtrisent ce qui est mesuré et ceux qui ne maîtrisent pas les notions, habiletés ou compétences évaluées; il s'agit de connaître certaines caractéristiques de ces deux groupes d'élèves. Des indices statistiques fournissent des renseignements sur la capacité des items à bien identifier les items dits « discriminants ».
- **La cohérence** Comme on le sait, une épreuve est constituée de plusieurs items ou questions. Une épreuve devrait mesurer un même aspect, une même discipline, une même dimension; on peut parler ici d'unidimensionnalité d'un instrument d'évaluation. Ainsi, on peut

s'interroger sur la cohérence qui devrait s'établir entre les différents renseignements recueillis. Il s'agit de la cohérence interne de l'instrument.

Les activités de ce séminaire ont permis aux participants de réaliser des compilations statistiques utilisant deux logiciels spécialisés : XCalibre² et EduStat.

Dans un deuxième temps, il y a eu présentation de la technique utilisant la **Théorie des Réponses aux Items** (TRI). Il s'agit ici d'un ensemble de procédés statistiques dont le but est d'évaluer la qualité d'un instrument de mesure et des items qui le composent. Comme l'objectif poursuivi par cette forme d'analyse est la généralisation à l'ensemble de la population possiblement concernée par l'épreuve examinée, il est nécessaire de disposer d'information provenant d'un nombre suffisant de personnes soumises à l'épreuve analysée. Cette approche d'analyse fait gagner en précision mais aussi en complexité. Trois paramètres peuvent être retenus au moment de la réalisation des compilations utilisant cette technique :

- La **discrimination**, c'est-à-dire le pouvoir de l'item de bien identifier les élèves qui maîtrisent ce qui est mesuré et ceux qui ne maîtrisent pas les notions, habiletés ou compétences évaluées (ce paramètre est identifié par la lettre «a»).
- La **difficulté** de l'item, c'est-à-dire la plus ou moins grande facilité de l'item pour les élèves visés par l'épreuve (il s'agit du paramètre «b»).
- La **pseudo chance**, c'est-à-dire la possibilité pour l'élève d'obtenir une réponse correcte au hasard sans posséder réellement l'habileté requise (c'est le paramètre «c»).

Dans le cadre de ce séminaire, le logiciel XCalibre a été utilisé pour illustrer les compilations réalisées par rapport à l'analyse TRI. Le logiciel EduStat, pour sa part, permet la préparation des fichiers (les données et les paramètres de calcul) nécessaires au bon fonctionnement de XCalibre.

Au cours du séminaire, il y a aussi eu présentation de techniques permettant de réaliser des analyses sur le fonctionnement différencié des items. Il est, en effet, important de s'assurer que certaines questions faisant partie d'un instrument

² Ce logiciel est développé et distribué par [Assessment Systems](#).

d'évaluation ne favorisent pas un sous-groupe d'élèves au détriment des autres. Il s'agit de l'examen de biais qui peuvent toucher plusieurs dimensions socio-économiques ou culturelles. On peut vouloir examiner s'il y a présence de biais par rapport à deux sous-groupes d'élèves (par exemple, les garçons et les filles; le milieu urbain et le milieu rural) ou bien plusieurs (par exemple, les régions administratives). Les techniques d'analyses seront différentes pour chacune de ces situations. Les items identifiés comme « biaisés » à partir de la valeur des indices calculés devraient être retranchés de l'épreuve ou bien modifiés. Le logiciel EduStat permet de réaliser des compilations permettant d'identifier de possibles items biaisés. Les participants ont pu réaliser de telles compilations avec les données disponibles.

Adresse courriel : info@mcpythagore.com