

## Mesure et évaluation en éducation



# ÉLABORATION D'UN EXAMEN DE RENDEMENT SCOLAIRE : TECHNIQUES ET PROCÉDURES. Mamdouh, Dawoud, Montréal : Éditions nouvelles, 129 pages, 1995

Serge P. Séguin

Volume 18, numéro 2, 1995

URI : <https://id.erudit.org/iderudit/1092281ar>

DOI : <https://doi.org/10.7202/1092281ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

ADMEE-Canada - Université Laval

ISSN

0823-3993 (imprimé)

2368-2000 (numérique)

[Découvrir la revue](#)

Citer ce compte rendu

Séguin, S. P. (1995). Compte rendu de [ÉLABORATION D'UN EXAMEN DE RENDEMENT SCOLAIRE : TECHNIQUES ET PROCÉDURES. Mamdouh, Dawoud, Montréal : Éditions nouvelles, 129 pages, 1995]. *Mesure et évaluation en éducation*, 18(2), 111–114. <https://doi.org/10.7202/1092281ar>

**ÉLABORATION D'UN EXAMEN DE RENDEMENT SCOLAIRE:  
TECHNIQUES ET PROCÉDURES**

Mamdouh, Dawoud,

Montréal : Éditions nouvelles, 129 pages, 1995.

En avant-propos (p. 9), l'auteur présente son volume comme contenant « des notions très utiles qui ouvrent la voie au futur maître, ...[et constituant] un instrument de travail précieux pour les enseignants de tous niveaux ». Il oriente ses propos en affirmant que l'enseignant doit « faire savoir combien d'objectifs ont été atteints si l'examen est critérié et, si le contexte est normatif, où se situe l'élève par rapport à sa classe ».

Pour [tenter d'] y parvenir, l'auteur propose un découpage en cinq chapitres: 1) les différentes étapes d'élaboration d'un examen; 2) les différents types de questions; 3) la validité, 4) la fidélité; 5) la présentation des résultats.

Cet ouvrage s'adresse donc à des enseignants ou à des étudiants peu ou pas initiés aux principes, techniques et procédures qui sont ou devraient être en vigueur en évaluation des apprentissages scolaires. D'une façon générale, ces peu ou pas initiés sont heureux de lire le volume de Dawoud : style clair, concis, organisé, avec juste ce qu'il faut de dose d'exemples pratiques et d'explications conceptuelles. Malheureusement, ces lecteurs sont, malgré eux, conduits à des compréhensions ou conceptions souvent insuffisantes ou inadéquates, voire erronées.

Nous sommes devant un texte qui, avec un discours dont l'apparente qualité pédagogique séduit, dissimule des embûches didactiques que nous ne pouvons passer sous silence, par suite de l'expérience que nous avons faite avec nos étudiants, enseignants et futurs enseignants.

Nous avons relevé trois types de causes à ces constatations : dans certains cas, l'auteur néglige simplement de faire des liens d'intégration qui seraient nécessaires à la compréhension de ce qui est présenté ; ailleurs, l'auteur appuie certains principes sur des données trop partielles, voire insuffisantes ; enfin, certains énoncés sont carrément erronés.

Nous ne choisirons ici que les principaux exemples justifiant notre jugement sévère. Les trois premiers commentaires font ressortir le manque de pertinence ou d'utilité de trois des cinq chapitres du volume. Les autres exemples font ressortir que les chapitres 3 et 4, malgré leur pertinence, comportent plusieurs embûches pédagogiques ou scientifiques.

Au chapitre premier, le seul modèle proposé de table de spécifications n'est approprié que pour un examen sommatif sur un contenu dont les objectifs opérationnels sont absents (ou comme si de tels objectifs n'étaient pas formulés). Une telle table ne convient pas à l'objectif, que l'auteur énonce pourtant à plusieurs endroits, voulant que l'évaluation des apprentissages doive permettre de savoir quels objectifs d'apprentissage les élèves maîtrisent. La fin du chapitre présente l'analyse d'items d'une façon difficile à comprendre pour le non-initié, et dans un contexte qui n'est pas celui des enseignants.

Au chapitre 2, plusieurs des arguments en faveur de questions à choix multiples, simples ou complexes, à quatre ou cinq choix, contredisent les qualités au nom desquelles l'auteur prône, dix pages auparavant, l'utilisation de questions à deux choix. L'auteur néglige de dire que ces dernières ne peuvent être tolérées qu'à deux conditions : le nombre de questions par objectif doit être élevé si l'on veut pouvoir conclure avec un minimum de crédibilité ; il est impossible de faire autrement. Or, justement, même lorsqu'un objectif porte sur une classification binaire ou dichotomique, il est toujours possible de composer des questions à quatre ou cinq choix dont chacun exige du répondant qu'il le classe selon la dichotomie proposée.

Le chapitre 5 a le mérite de bien expliquer en détail moyenne, médiane, mode, selon différents types de barèmes de fréquences. Mais ce dont les enseignants ont besoin, c'est aussi des rangs cinquièmes et des rangs centiles, et non des centiles. Et pour qui veut comprendre les fameux scores- $x$ , c'est aussi le silence. Aucune interprétation normative autre que le rang n'est couverte... pas plus d'ailleurs qu'aucune interprétation critériée. Le titre de ce chapitre ne parle que de la présentation des résultats ; ce qui est juste. Mais c'est le but même que l'auteur énonce qui se trouve alors contredit : rien dans cet ouvrage ne permet à l'enseignant d'apprendre comme procéder pour « faire savoir combien d'objectifs ont été atteints », ni « où se situe l'élève par rapport à sa classe ».

Au total, jusqu'à maintenant, le volume de Dawoud apparaît doublement inutile. Non seulement les premier et dernier chapitres ne permettent pas de répondre adéquatement aux objectifs ciblés par l'auteur, mais encore les contenus de ces chapitres et du deuxième sont déjà plus adéquatement couverts par les ouvrages de Tousignant et Morissette (1990) et de Morissette (1993). C'est donc par ses deux autres chapitres que le volume de Dawoud est susceptible d'une contribution attendue : il y a en effet place, dans la documentation scientifique de langue française, pour une initiation

scientifique aux concepts et techniques de fidélité et de validité des examens. Mais il s'agit d'un défi qu'il n'est pas facile de relever.

Au chapitre 3 traitant de validité, l'auteur situe la corrélation item-test (question-total) en validité de contenu, mais l'indice de discrimination, l'usuel D exprimant la différence entre les proportions de réussites dans les groupes extrêmes, en validité congruente prédictive [sic]. Ces deux techniques appartiennent pourtant toutes deux à l'analyse du pouvoir de discrimination des items, laquelle constitue une vérification empirique de la validité de contenu. Il s'agit ici d'un parfait exemple de confusion conceptuelle chez l'auteur, conduisant à des erreurs d'apprentissage chez le lecteur : principes et techniques d'analyse d'items ne sont pas intégrés.

De même en est-il de l'autre exemple donné par l'auteur de la validité qu'il appelle congruente prédictive. La technique admissible ici consiste à comparer la performance au critère Y de deux groupes extrêmes sur le test X. L'auteur fait donc une double erreur : il décrit, sans le dire, et faussement, le processus de l'indice D ; il omet de souligner que le procédé de validité décrit (ou celui qu'il aurait dû décrire) produit un coefficient dit de concordance et peut être utilisé alternativement au coefficient de corrélation dans tout cas où celui-ci s'applique.

Plus loin, en énumérant sept critères pour apprécier la validité hypothético-déductive, l'auteur néglige de faire les liens qui s'imposent avec les autres types de validité ou avec les techniques préalablement présentées. Ainsi, le « critère » [sic] 2 et 4 sont, en fait, des techniques d'analyse d'items.

Quant au coefficient alpha de Cronbach, dont l'écriture de la formule est typographiquement incorrecte, l'auteur le présente faussement comme indicateur de validité de construit. En outre, au chapitre 4 traitant de fidélité, l'auteur néglige de faire les liens depuis 30 ans reconnus entre ce coefficient et tout autre coefficient de fidélité par homogénéité ou consistance interne.

Dans l'ensemble du chapitre sur la validité, on ne tient pas compte de l'existence de banques d'items ; des exposés formels d'autres types de validité sont éloignés du contexte de l'enseignement ; l'utilisation de références dépassées et sans regard critique amène à conclure faussement qu'augmenter le nombre d'items en général, ou ceux de difficulté moyenne en particulier, augmenterait automatiquement la validité ou la discrimination, respectivement.

Dans l'écriture d'un volume de ce type, l'auteur doit savoir apprécier l'importance de traiter tels ou tels aspects, et les traiter selon un principe intégrateur ou un fil conducteur. Dans l'ouvrage de Dawoud, on assiste plutôt à un parachutage d'objets sans relation, sans intégration, sans fil conducteur ; ou avec un prétendu fil conducteur qui n'est pas respecté, ce qui a le même effet.

Les besoins de traitements conceptuels et méthodologiques, en langue française, de la validité et de la fidélité et des techniques qui en émergent, ainsi que des principes et techniques d'analyse d'items, ne sont donc pas adéquatement comblés par le volume de Dawoud.

### RÉFÉRENCES

- MORISSETTE, D. (1993). Les examens de rendement scolaire (3<sup>e</sup> éd.) Québec : Presses de l'Université Laval.
- TOUSIGNANT, R. & MORISSETTE, D (1990). Les principes de la mesure et de l'évaluation des apprentissages (2<sup>e</sup> éd.). Boucherville (Québec) : Gaétan Morin.

Serge P. Séguin  
Université du Québec à Montréal