



Examen de divers tests:
Outils d'évaluation en milieu de
travail

LIRE TOUS LES JOURS!



Table des matières

Introduction

Partie 1 Dimensions du contrôle des connaissances

L'évaluation en situation d'apprentissage

Quelles sont les compétences essentielles dans le monde d'aujourd'hui?

Types de test

Partie 2 Examen technique détaillé de divers tests

Tests of Adult Basic Education (TABE) Work-Related Foundation Skills

Contenu

Scores

Processus d'élaboration

Tests of Applied Literacy Skills

Contenu

Scores

Processus d'élaboration

Caractéristiques relatives aux échelles et aux tests

Basic English Skills Test

Contenu

Scores

Processus d'élaboration

Canadian Adult Achievement Test

Contenu

Scores

Processus d'élaboration

Measurement, Reading and Arithmetic

Contenu

Scores

Processus d'élaboration

Wide Range Achievement Test

Contenu

Scores

Processus d'élaboration

Metric Skills Assessment

Contenu

Scores

Processus d'élaboration

INTRODUCTION

L'évaluation des participants aux programmes d'alphabétisation en milieu de travail est toujours l'objet de vifs débats dans le domaine de l'éducation des adultes. C'est une question qui peut être abordée de divers points de vue, sous l'angle tantôt des composantes essentielles de tout test normalisé tantôt des moyens d'obtenir l'information nécessaire à la prise de décisions. Dans le document qui suit, nous nous intéressons à un aspect de l'évaluation, à savoir le contrôle des connaissances. Nous procédons à un examen technique détaillé de sept outils d'évaluation employés dans le domaine de l'éducation en milieu de travail. Nous examinons chaque instrument au regard du contenu, des scores et du processus d'élaboration.

Pour bien situer le sujet dans son contexte, nous commençons par examiner certaines des dimensions du contrôle des connaissances. Cet examen préalable pourra servir de point de départ pour aider les praticiens à choisir les tests à utiliser quand il faut évaluer les participants à un programme d'éducation en milieu de travail. Le but du document qui suit est de montrer, non pas que certains tests sont supérieurs à d'autres, mais que les formateurs en milieu de travail doivent trouver le test qui convient le mieux aux besoins de leurs apprenants et de leurs programmes respectifs.

PARTIE 1

DIMENSIONS DU CONTRÔLE DES CONNAISSANCES

L'évaluation en situation d'apprentissage

L'évaluation dans le domaine de l'éducation en milieu de travail peut être considérée comme une carte routière qui aide les yens à arriver à une destination en particulier. La carte ou le plan d'apprentissage guide le stagiaire vers une destination qui correspond aux buts et objectifs du programme. Les yens se servent souvent de cartes routières pour déterminer la route à prendre pour arriver à leur destination; certains veulent la route la plus rapide, tandis que d'autres préfèrent une route plus pittoresque qui comprend des points d'intérêt. Il en est de même pour l'évaluation, qui devrait aider l'apprenant à avancer vers ses objectifs et lui indiquer combien de temps il lui faudra pour les atteindre. Dans beaucoup de programmes d'éducation offerts actuellement en milieu de travail, l'évaluation est utilisée pour répartir les apprenants selon leur niveau, pour leur prescrire des plans d'apprentissage propres à leurs besoins, pour contrôler leurs progrès et pour vérifier la réalisation des objectifs. Beaucoup de praticiens et d'universitaires estiment que l'évaluation bien faite est un élément essentiel du plan d'apprentissage au même titre que le programme d'études et l'enseignement. Le plan d'apprentissage est constitué de diverses stratégies et techniques d'enseignement, d'évaluation et de renforcement de l'apprentissage. Aucune méthode d'enseignement ne peut être efficace pour tous les apprenants dans toutes les circonstances; de même, aucun instrument ni

procédé d'évaluation ne permet de fournir de [l'information de qualité qui réponde à tous les besoins.

De manière générale, l'évaluation dans le domaine plus vaste de l'alphabétisation des adultes se fait selon trois grandes approches, qui peuvent être utiles pour mieux comprendre l'évaluation dans le domaine de l'alphabétisation en milieu de travail. Chaque approche repose sur certaines hypothèses et comporte certaines limites. Selon l'approche traditionnelle, l'alphabétisation fonctionnelle vise l'obtention aux tests de rendement d'un score correspondant à un niveau scolaire en particulier. Ces tests permettent d'évaluer le pourcentage de divers groupes dont le rendement est égal ou supérieur à un niveau scolaire particulier en lecture. Ceux dont le rendement est égal ou supérieur au niveau en question sont présumés avoir les compétences voulues pour accomplir des tâches ou se servir de documents correspondant à un niveau de difficulté comparable.

La deuxième approche se fonde sur [l'acquisition de compétences. Des tests comportant des tâches élaborées à partir de matériel de type non scolaire sont administrés à des échantillons nationaux. L'analyse des résultats obtenus permet d'évaluer la proportion de la population adulte qui devrait pouvoir accomplir avec succès les tâches en question et de déterminer dans quelle mesure divergent les caractéristiques, comme le niveau d'instruction, l'origine raciale, l'origine ethnique, le sexe et le niveau de revenu, sont liées aux taux estimatifs de réussite.

Selon l'approche profil, l'alphabétisation est considérée, non pas simplement comme [l'acquisition d'un ensemble de compétences isolées liées à la lecture et à l'écriture, mais davantage comme [l'application de ces compétences à des fins précises et

dans des contextes précis. Étant donné la diversité des contextes d'application, au foyer, au travail et en société, l'alphabétisation est perçue, non pas comme une norme unique, mais comme un ensemble de pratiques utilisées dans divers contextes. Cette approche comprend trois aspects distincts de l'alphabétisation: les textes, les documents et les nombres.

Le lecteur constatera que chacune de ces trois façons d'aborder l'évaluation se trouvent représentées dans la partie examen technique détaillé de divers tests.

Quelles sont les compétences essentielles dans le monde d'aujourd'hui?

Les divers types de compétences nécessaires dans le monde du travail ont une autre dimension de l'évaluation et du contrôle des connaissances. Pour choisir un instrument d'évaluation, il faut notamment se demander si les compétences qu'il évalue sont du même type que les compétences devant être appliquées dans l'exercice des fonctions actuelles ou futures. C'est là un domaine d'enquête qui suscite l'intérêt depuis déjà un certain temps. Parmi les travaux de recherche en cours, il convient de signaler le Projet de recherche sur les compétences essentielles de Développement des ressources humaines Canada, qui vise à déterminer ce que font les travailleurs de divers secteurs. Les chercheurs ont interviewé quelque 3 000 travailleurs d'une multitude de secteurs au Canada afin de répertorier les compétences essentielles effectivement nécessaires pour les occupations de premier échelon exigeant des compétences minimales. L'étude devait se solder par un profil décrivant les compétences essentielles pour chacune de ces occupations dans le contexte canadien. Les profils ainsi obtenus comprennent aussi bien les compétences effectivement utilisées au travail telles qu'elles sont décrites par les travailleurs en possession que les compétences de base traditionnelles, soit la capacité de raisonnement, la capacité de travailler en équipe et les compétences nécessaires à la formation permanente. Quand on a une idée générale de ces différentes catégories de compétences, on peut mieux choisir le type d'instrument d'évaluation à utiliser. Il convient de signaler ici qu'aucun des sept tests examinés dans la suite du document n'est destiné à mesurer toutes les catégories de compétences décrites dans les paragraphes qui suivent.

La lecture de textes, la recherche documentaire, la rédaction, le calcul et la communication valent au nombre des compétences de base traditionnelles. Voici des exemples de compétences qu'on retrouve dans certaines de ces catégories. Dans la catégorie lecture de textes, certaines tâches typiques ou plus complexes sont évaluées selon une échelle qui va de la lecture de textes relativement courts jusqu'à l'interprétation de textes denses et complexes qui comporte un niveau de difficulté plus grand, en passant par la recherche d'un élément d'information en particulier ou la compréhension d'instructions écrites simples.

La catégorie recherche documentaire comprend des tâches typiques ou plus complexes en lecture. Ces tâches sont évaluées selon le niveau de difficulté et vont de la lecture d'affiches, d'étiquettes et d'autres textes semblables où il suffit de pouvoir retrouver des mots clés jusqu'aux déductions fort abstraites faites à partir de textes et à l'utilisation de connaissances spécialisées. Dans la catégorie rédaction, on retrouve des tâches typiques ou plus complexes selon le cas qui sont aussi évaluées, avec au bas de l'échelle les compétences comme la rédaction de textes très courts à insérer dans un formulaire et au haut les tâches plus complexes comme la rédaction de textes sur une vaste gamme de sujets en vue de persuader quelqu'un ou de présenter une analyse, une comparaison, une évaluation ou une critique.

Dans la catégorie calcul, les tâches sont réparties par domaine d'application, selon qu'il s'agit d'opérations financières simples, de l'établissement d'horaires ou de budgets, de mesures ou d'analyse de données. La capacité de calcul englobe aussi des compétences plus précises ayant trait aux entiers, aux fractions, aux proportions, à la conversion de

mesures et à la géométrie. La capacité de communication est évaluée selon le type de communication et le but visé, comme d'écouter, de parler, de traiter avec les camarades de travail, de traiter avec les superviseurs, de traiter avec les clients, de participer à des discussions de groupe, de présenter de l'information à des groupes nombreux et d'animer des discussions structurées.

La deuxième grande catégorie de compétences concerne la capacité de raisonnement, qui comporte diverses sous-catégories. La première sous-catégorie est celle des compétences liées à la résolution de problèmes, qui sont aussi évaluées selon divers facteurs, comme la complexité du problème, la difficulté de le définir, la difficulté de définir les étapes permettant d'en arriver à une solution et la difficulté d'évaluer la situation. Vient ensuite la prise de décisions qui est liée à la résolution de problèmes. Encore là, les tâches typiques ou plus complexes sont cotées selon une échelle qui tient compte de facteurs comme la présence d'un ensemble de procédures ou d'un arbre de décision dont on peut s'inspirer, la conséquence des erreurs et le degré de jugement nécessaire pour prendre une décision acceptable. Toujours sous la rubrique des compétences liées à la capacité de raisonnement, on trouve la sous-catégorie des compétences permettant de planifier et d'organiser son travail. Ces compétences sont cotées selon une échelle qui tient compte de facteurs comme la diversité des fonctions à accomplir et la mesure dans laquelle le plan de travail de l'employé doit s'intégrer aux plans de travail d'autres employés. Viennent ensuite les compétences liées à la mémoire. Ces compétences sont évaluées selon le type de mémoire auxquelles elles font appel: se rappeler une certaine information pendant un court laps de temps, se rappeler des

événements précis et se rappeler de ses informations propres à son travail pour s'en servir à long terme.

Dans la troisième grande catégorie, la capacité à travailler avec d'autres, les compétences sont réparties selon le contexte dans lequel le travail se fait, comme de travailler seul, de façon indépendante, avec un partenaire ou un assistant ou comme membre d'une équipe, et selon qu'il s'agit de jouer un rôle de supervision ou de leadership, comme d'orienter les nouveaux employés, d'attribuer des tâches à d'autres travailleurs, de choisir des entrepreneurs ou des fournisseurs ou de participer à des discussions structurées sur la façon dont le travail doit se faire. Les compétences informatiques constituent la catégorie suivante. L'utilisation d'ordinateurs est évaluée selon une échelle de un à cinq points, le plus bas niveau étant celui qui consiste à se servir d'un ordinateur sans avoir besoin de connaître le logiciel requis et le niveau le plus élevé étant celui qui consiste à évaluer les besoins et à concevoir des systèmes informatiques.

En résumé donc, les compétences essentielles sont celles qu'on retrouve dans presque tous les types d'occupation, même si la forme ou le niveau de complexité peut varier selon le type d'occupation. Ces compétences sont essentielles à l'acquisition des compétences plus précises qui permettent d'accomplir les tâches propres à un type de travail en particulier.

Types de tests

Beaucoup de praticiens trouvent difficile de prendre des décisions judicieuses quant aux outils d'évaluation à utiliser parce qu'ils ont du mal à s'y retrouver dans la terminologie technique des tests. Bien que le sujet ait déjà été abordé dans d'autres documents, la section qui suit décrit certains des termes du vocabulaire des tests. Normalisation, normes, norme absolue et tests axés sur les compétences vont autant de termes que le lecteur rencontrera dans la partie examen technique détaillée de divers tests. Test normalisé s'entend d'un test qui est administré dans des conditions normalisées et qui permet de faire des déductions au sujet de la capacité de l'apprenant à partir d'un échantillon de performances étalonnées. Ce qui distingue le test normalisé du test informel, c'est qu'il est administré dans des conditions uniformes. Ainsi, le test de rendement normalisé comporte les mêmes textes et les apprenants doivent suivre les mêmes procédures, répondre aux mêmes types de questions et respecter les mêmes limites de temps.

Il existe deux types de tests normalisés: le test normatif et le test critériel. Beaucoup de tests normalisés vont conçus pour permettre d'interpréter le score de l'apprenant par rapport ou par référence aux scores obtenus par d'autres sujets qui ont fait le test. C'est ce qu'on appelle le test normatif. Le score qu'obtient le sujet est interprété en le comparant aux performances étalonnées du groupe de référence.

L'autre type de test normalisé est le test critériel. L'utilisation de ce type de test va de pair avec l'élaboration de programmes d'enseignement individualisés où chacun apprend à son rythme. Selon cette approche, le champ des connaissances ou des

compétences à acquérir est soigneusement défini. Le score de chaque apprenant est déterminé par rapport à une norme ou un critère de performance absolu. Les programmes de ce genre exigent d'établir des objectifs d'apprentissage précis et de prévoir des unités d'enseignement, souvent appelées modules, pour enseigner les divers soul-ensembles de connaissances et de compétences définis dans les objectifs d'apprentissage.

L'évaluation peut aussi se faire à partir de tests axés sur les compétences. La première étape de l'élaboration d'un test de ce genre consiste à établir les objectifs clés à atteindre ou les principales compétences à acquérir. On arrête ensuite les éléments du test de manière à déterminer si l'apprenant possède déjà les compétences voulues ou s'il a besoin d'un programme d'enseignement pour acquérir certaines compétences. Enfin, on dresse la liste des divers matériels pédagogiques offerts sur le marché et qui comportent une gamme variée d'exercices visant l'acquisition de chacune des compétences de façon que les enseignants puissent choisir ce qui convient à leurs apprenants. Selon cette approche, l'enseignement se fait en quelque sorte en fonction du test, même si le contenu comme tel du test ne figure pas nécessairement au programme d'études afin d'éviter que l'enseignement ne soit directement lié aux divers éléments du test. Les programmes axés sur les compétences exigent donc qu'on définisse d'abord les compétences à acquérir et qu'on structure le programme d'études afin d'aider l'apprenant à acquérir ces compétences.

PARTIE 2
TESTS OF ADULT BASIC EDUCATION (TABE)
(tests d'éducation de base pour adultes)

Work-Related Foundation Skills
CTB MacMillen-MacGraw Hill,
20, chemin Ranch,
Monterey (California), 93940-5703,
Tél. :1-800-538-9547

Contenu

La Work-Related Foundation Skills (WRFS) est une collection de tests normatifs visant à déterminer dans quelle mesure le sujet possède les compétences de base en lecture, en calcul et en activités langagières qu'on retrouve dans la majorité des programmes d'éducation de base pour adultes. À l'aide d'un contenu exprimé dans un langage judicieusement choisi pour représenter le contexte du milieu de travail, ces tests déterminent dans quelle mesure le sujet possède les compétences voulues pour fonctionner au travail et en société et quels sont ses besoins à cet égard.

La WRFS comprend des tests dans quatre domaines: santé, métiers professionnels techniques, commerce-bureau et général. Les questions, les graphiques et les passages qui composent chaque test sont présentés dans un contexte apparenté à ces différents milieux de travail. Les quatre tests de la collection WRFS mesurent des compétences comparables. Comme leur contenu est comparable, les quatre tests peuvent être utilisés efficacement soit avant soit après la formation.

Chacun des quatre tests de la WRFS comprend les sous-tests suivants: lecture, calcul arithmétique, mathématiques appliquées et activités langagières. Chaque sous-test dure une trentaine de minutes. Le sous-test de lecture porte sur la compréhension du

processus de lecture ainsi que sur la capacité à trouver le sens d'un texte. À l'aide de passages écrits, de graphiques et de références, le soul-test évalue la capacité du sujet à reconnaître des mots dans un contexte de travail, à se rappeler des informations et à retrouver les idées principales. Les questions visent à mesurer la capacité de lecture fonctionnelle de diagrammes, de tableaux, de formulaires et d'horaires.

Le sous-test de calcul arithmétique vise à déterminer dans quelle mesure le sujet possède les compétences de base voulues pour suivre avec succès un programme d'enseignement en mathématiques, dans quelle mesure il comprend les différentes opérations arithmétiques, addition, soustraction, multiplication et division, avec des entiers, des décimales et des fractions. Il vise aussi à mesurer sa compréhension des opérations financières simples, des entiers et des pourcentages. Le soul-test de mathématiques appliquées porte principalement sur la formation de concepts, le raisonnement et la résolution de problèmes en contexte. Il mesure la capacité du sujet à appliquer des concepts mathématiques liés à la rémunération, au calcul en contexte, à la théorie des nombres, aux mesures, à l'interprétation de données, à la géométrie et à la pré-algèbre. Il évalue aussi la capacité de raisonnement nécessaire à la résolution de problèmes pratiques.

Le soul-test de langue mesure les compétences du sujet en ce qui concerne les règles de l'usage, la formation de phrases, le développement de paragraphes et les conventions d'écriture comme l'utilisation des majuscules et la ponctuation. Le soul-test porte sur les compétences essentielles pour être considéré comme pleinement alphabétisé et met l'accent sur le lien entre langage, lecture et écriture. Il s'intéresse aussi à

l'utilisation correcte du langage dans le milieu de travail. Le test présente une multitude de situations de travail authentiques en guise de stimuli.

Scores

La collection WRFS fournit des scores normatifs sous forme d'échelles de scores, d'équivalents de niveau scolaire et de rangs centiles. Il fournit aussi des informations critiques relatives aux objectifs incorporés aux tests. Les sous-tests et les objectifs de la collection WRFS ont été établis à partir d'informations tirées des programmes d'éducation des adultes de divers États et après examen des résultats et des profils de compétences de programmes d'éducation des adultes à l'échelle des États-Unis. L'échantillon devant servir d'étalon pour les tests de la WRTS comprenait environ 1 700 apprenants adultes sélectionnés à l'échelle du pays en 1994. Les tests TABE de la collection WRTS de même que les formulaires 7 et 8 ont été administrés à chacun des participants.

Processus d'élaboration

Les tests de la collection VWRFS ont été élaborés selon un processus typique comprenant les sept étapes suivantes: conception de l'évaluation, rédaction, examen des items, essai, analyse des items, choix du test et élaboration des normes.

Fiabilité. Les tests de la VWRFS obtiennent un coefficient Kuder-Richardson formule 20 variant entre 0,82 et 0,91. Le coefficient KR 20 mesure la validité des scores. Plus il est élevé, plus les différents sous-tests témoignent d'une cohérence interne élevée. Par cohérence interne, il faut entendre que chaque sous-test mesure dans son ensemble une

seule compétence, comme la capacité de lecture, plutôt qu'une combinaison de plusieurs compétences.

Validité. La validité du contenu des tests de rendement tient à la concordance entre le contenu du test et le contenu du programme d'enseignement. Pour assurer cette concordance, les responsables de l'élaboration des tests ont examiné une multitude de programmes d'études et ont rencontré des spécialistes en éducation afin de déterminer les objectifs de rendement commune aux divers programmes existants de même que les connaissances et les compétences sur lesquelles on insiste dans ces programmes.

L'information ainsi obtenue les a guidés dans toutes les étapes de la conception et de l'élaboration de cet outil d'évaluation des programmes d'éducation en milieu de travail.

Validité conceptuelle. La validité conceptuelle, qui englobe l'interprétation des scores et les déductions qui peuvent être faites à partir de ces scores, est un élément essentiel du processus de validation des outils d'évaluation utilisés en milieu de travail. Il existe d'abondantes preuves de la validité conceptuelle, preuves qui ont trait tant au contenu qu'aux critères. Ainsi, les outils d'évaluation de l'éducation en milieu de travail sont exhaustifs dans la mesure où ils comprennent des éléments qui reflètent les objectifs d'enseignement essentiels. En outre, les corrélations entre ces outils d'évaluation et entre les outils et les autres instruments de la famille des instruments d'évaluation TABE témoignent d'une validité convergente et divergente. Autrement dit, les tests qui sont conçus pour mesurer des compétences semblables ont une corrélation plus élevée que les tests conçus pour mesurer des compétences tout à fait différentes. Ainsi, le sous-test de langue de la WRFS est en étroite corrélation avec le sous-test de lecture, comme on

pouvait s'y attendre, tandis que la corrélation avec les sous-tests de mathématiques de la WRFS est moins élevée.

TESTS OF APPLIED LITERACY SKILLS

(tests mesurant le degré d'alphabétisation fonctionnelle)

**Educational Testing Service
Language Learning and Assessment
Chemin Rosedale, Arrêt postal 05-9,
Princeton (New Jersey) 08541,
Tél.: 1-800-551-1230**

Contenu

Les tests du degré d'alphabétisation fonctionnelle (TALS) mesurent trois aspects distincts et importants du degré d'alphabétisation: la lecture de textes, la recherche documentaire et le calcul. Il existe deux formulaires pour chacun des trois soul-tests. Les tâches relatives à la lecture de textes exigent les connaissances et les compétences nécessaires pour comprendre et utiliser l'information contenue dans divers textes, notamment des éditoriaux, des articles de journaux et des poèmes. Les deux formulaires pour le test de lecture de textes comprennent chacun 48 tâches. Les textes utilisés sont surtout de nature descriptive. Ils sont reproduits intégralement suivant la mise en pages et la typographie de la source initiale. Le soul-test évalue trois aspects de la lecture de textes: trouver de l'information, l'intégrer et la produire. Les tâches liées à chacun de ces trois aspects comportent des niveaux de difficulté divers, tout comme celles qui sont incluses dans les trois soul-tests. Une période de 20 minutes est accordée pour chacune des deux sections de chaque formulaire.

Les tâches liées à la recherche documentaire englobent les connaissances et les compétences nécessaires pour trouver et utiliser l'information contenue dans des demandes d'emploi, des formulaires de paye, des horaires de transport, des cartes routières, des tableaux, des index, etcetera. Les deux formulaires pour le test de recherche

documentaire comprennent chacun 52 tâches. L'accomplissement de beaucoup de ces tâches est nécessaire pour répondre aux exigences d'un emploi et gérer un ménage. Le test mesure la maîtrise des techniques nécessaires pour trouver de l'information dans divers documents complexes, comme des graphiques et des tableaux, et pour transférer cette information d'un document à un autre. Trois types de questions ou d'instructions, amenant le sujet à trouver de l'information, à la parcourir et à l'intégrer, sont utilisés pour évaluer les compétences.

Les tâches liées au calcul englobent les connaissances et les compétences nécessaires pour faire des opérations arithmétiques, soit isolément soit en ordre séquentiel, à partir de chiffres figurant dans des documents: par exemple, établir le solde d'un compte de chèque, fixer un itinéraire de voyage, remplir un formulaire de commande ou déterminer le montant de l'intérêt sur un prêt ou un placement. Pour le test de calcul, on peut choisir entre deux formulaires comportant chacun 46 tâches. Le sujet doit faire des additions, des soustractions, des multiplications et des divisions à partir de chiffres compris dans des documents qu'on retrouve dans la vie de tous les jours. Le degré de difficulté des tâches est fonction des opérations de calcul exigées, de la mesure dans laquelle les chiffres figurant dans les documents s'imposent à l'esprit et des déductions qui doivent être faites, le cas échéant, pour déterminer le type d'opérations nécessaires.

Scores

La première étape de l'interprétation des résultats du test consiste à calculer le score brut. Le score brut correspond au nombre total de bonnes réponses dans chaque section du formulaire, les réponses étant notées à partir du guide inclus dans le manuel

qui accompagne le test. La deuxième étape consiste à convertir les scores bruts en scores de performance à partir des tableaux qui se trouvent dans le manuel correspondant à chacun des formulaires des trois tests. Pour déterminer le score de performance du sujet, l'évaluateur choisit le tableau qui correspond au formulaire afin d'y trouver le nombre de bonnes réponses pour chacune des deux sections du test. Le chiffre marquant dans le tableau [l'intersection des rangées correspondant à la section visée est le score de performance. Ce score donne une estimation de la compétence qu'a démontrée le sujet dans chacun des domaines de compétences à évaluer.

Une fois qu'il a en main le score de performance, l'évaluateur se sert du cadre d'interprétation inclus dans le manuel, qui permet de repérer sur chacune des échelles d'alphabétisation le niveau de difficulté correspondant aux différentes tâches. Les échelles ont été conçues pour mesurer toute la gamme des degrés de difficulté de 0 à 500, mais la majorité des tâches se situent entre 200 et 400. Pour comprendre ce que chaque test vise à mesurer, les tâches ont été interprétées à la lumière des interactions entre les stimuli et les questions ou instructions qui permettent de les situer sur chacune des échelles de difficulté. Le cadre d'interprétation du test de lecture de textes porte sur la capacité à retrouver de l'information dans un texte, à l'intégrer et à produire de l'information. De même, le cadre d'interprétation du test de recherche documentaire porte sur la capacité à retrouver de l'information dans divers documents, à parcourir l'information et à l'intégrer. Les conditions et les niveaux de difficulté du test de calcul mettent l'accent sur les opérations d'addition, de soustraction, de multiplication et de division et sur des combinaisons de ces différentes opérations.

Pour l'apprenant, les tests mesurant le degré d'alphabétisation fonctionnelle peuvent être utiles à maints égards. Ils lui fournissent une estimation de sa capacité démontrée à appliquer diverses compétences en lecture, en écriture et en calcul à certains documents qui font partie de la vie de tous les jours, au travail, au foyer et en société. Les scores de performance sont une mesure efficace de la gamme de tâches que le sujet devrait pouvoir systématiquement accomplir avec succès. Les scores obtenus peuvent aussi indiquer quel serait le matériel pédagogique qui conviendrait le mieux aux besoins. En outre, les scores fournissent à l'apprenant des informations utiles pour planifier son programme d'études et lui permettent d'évaluer son évolution et ses progrès.

Processus d'élaboration

Les tâches de simulation ont été élaborées à partir du principe que la capacité démontrée à accomplir une tâche donnée reflète l'interaction entre la structure de l'élément stimulus, le contenu représenté et ce que le sujet doit faire de l'élément en question. Les tests ayant été élaborés à partir de la théorie de la réponse d'item (IRT), il est possible d'évaluer le niveau de difficulté d'un exercice donné relativement au degré de difficulté de tous les exercices visés et d'estimer aussi le niveau de rendement du sujet dans le domaine évalué. La création des tableaux de conversion qui permettent d'obtenir à partir du score brut le score gradué estimatif sur l'échelle IRT comportait quatre étapes: a) l'établissement d'échelles IRT et l'estimation de la répartition rationnelle des personnes ayant participé à l'enquête Young Adult Literacy Survey (enquête sur le degré d'alphabétisation des jeunes adultes) de 1985; b) l'établissement d'échelles IRT distinctes à partir des éléments des tests TALS; c) l'établissement de liens entre les échelles TALS

et les échelles Young Adult Literacy Survey au moyen d'échantillons de liaison; et d) la création de tableaux de conversion des scores obtenus aux TALS.

Caractéristiques relatives aux échelles et aux tests

Probabilités de réponse. L'application du modèle IRT a notamment pour avantage de fournir de l'information sur les différents éléments des questionnaires qui permet de décrire les niveaux de rendement correspondent à chaque échelon de chaque échelle. Il devient possible de ce fait de déterminer les tâches que le sujet dont le score se situe à un niveau donné pourrait fort probablement accomplir avec succès, alors que le sujet dont le score se situerait à l'échelon juste en-dessous aurait des probabilités de succès bien moins élevées.

Relations inter-échelles. Les corrélations estimatives entre les trois tests mesurant le degré d'alphabétisation étaient respectivement de 0,59 (lecture de textes et recherche documentaire), de 0,60 (lecture de textes et calcul) et de 0,62 (recherche documentaire et calcul).

Rendement différentiel des groupes. Les stimuli retenus pour les tests ont été délibérément choisis pour refléter une vaste gamme d'intérêts et de contextes qu'on retrouve chez les adultes de façon qu'aucun groupe n'ait d'avantage par rapport à l'autre en raison de connaissances professionnelles très spécialisées. Par ailleurs, les questions ont été élaborées de manière à faire appel aux compétences liées aux divers types et niveaux de traitement correspondent aux trois domaines d'alphabétisation que l'on retrouve dans la vie de tous les jours.

BASIC ENGLISH SKILLS TEST

(test de compétences fondamentales en anglais)

**Centre for Applied Linguistics,
1118 nord-ouest, 22^e Rue,
Washington (DC) 20037,
Tél.: 202-429-9292**

Contenu

Le test de compétences fondamentales en anglais (BEST) a été conçu pour mesurer la capacité, de compréhension auditive, d'expression orale, de lecture et d'écriture correspondant au niveau fondamental et au niveau intermédiaire. Il se compose de deux volets, un volet entrevue orale et un volet degré d'alphabétisation, qui permettent de déterminer dans quelle mesure le sujet possède les compétences langagières fonctionnelles de base.

Le volet entrevue orale comprend 49 questions, réparties entre les domaines suivants: renseignements personnels, caractéristiques du foyer, heures, directions, achats et opérations financières simples, affiches, cabinet du médecin, accidents et formulaires d'emplois et de données. Le sujet doit aussi lire un passage et rédiger un texte. Ces deux tâches permettent de sélectionner les sujets à qui il y aurait peut-être lieu d'administrer la partie du test mesurant le degré d'alphabétisation. Toutes les questions simulent diverses situations réelles de compréhension auditive et d'expression orale. Cette partie du test consiste en un entretien personnel d'une quinzaine de minutes.

Le volet degré d'alphabétisation comprend une cinquantaine d'items classés selon les rubriques suivantes: renseignements personnels, calendrier, étiquettes apposées sur les aliments et les vêtements, chèque de foyer, enveloppes, annuaire téléphonique, horaire de

train, affiches, lecture d'un passage et rédaction d'un texte. Cette partie du test peut être administrée à un ou à plusieurs sujets et dure environ une heure.

Scores

Chacun des items du volet entrevue vale du BEST est conçu pour mesurer les compétences dans un des quatre domaines suivants: compréhension auditive, communication, maîtrise de la langue et lecture et écriture. Dans le guide d'accompagnement, chacun de ces quatre domaines est représenté par un symbole. Le guide contient une échelle de scores pour chaque item compris dans chaque domaine de compétences de même qu'un résumé des règles de notation pour l'entrevue orale. On trouve dans le guide une courte description et des exemples qui correspondent à chacun des chiffres de l'échelle, qui va de zéro à trois. Une fois qu'il a fait le total des scores à l'aide d'une feuille de notation, l'évaluateur détermine le score total pour l'entrevue vale. Les sujets qui obtiennent un score de cinq ou plus peuvent être appelés à faire le test mesurant le degré d'alphabétisation.

Une feuille de notation du degré d'alphabétisation est utilisée pour ce volet du test. Un score de un ou de zéro est donné pour chacune des onze sections du test. Le guide d'accompagnement comprend des exemples et des instructions sur la façon de noter. Le score total pour l'ensemble des questions est inscrit sur la feuille de notation. Suivant l'approche axée sur les compétences, sept niveaux de rendement ont été établis pour chacun des volets du test, chaque niveau correspondant à une échelle de scores différente.

L'information obtenue grâce aux deux volets du test peut être utilisée à des fins diverges. Les tests de classement ou d'admission évaluent la nature et l'étendue des

compétences langagières des étudiants inscrits à un cours d'anglais. En classant les étudiants selon le score obtenu au test BEST, on peut obtenir des classes plus ou moins homogènes. Le test peut également servir à mesurer les progrès des étudiants dans l'acquisition des compétences langagières fonctionnelles. Il peut aussi déterminer les progrès réalisés dans un cours en particulier à condition que le cours soit conçu pour enseigner les compétences langagières nécessaires pour survivre en société ou pour occuper un emploi. Le test peut également être utilisé à des fins diagnostiques. Le test BEST est conçu pour refléter le plus fidèlement possible les situations dans lesquelles les étudiants seront appelés à utiliser leurs compétences langagières pour pouvoir vivre de façon autonome. En examinant la performance de l'étudiant relativement à chacun des items du test, on obtient des informations importantes quant à sa capacité à accomplir des tâches langagières données. Enfin, le test peut être utilisé à des fins de sélection, notamment pour déterminer si l'étudiant est en mesure de suivre un programme d'études ou de formation donné.

Processus d'élaboration

Les spécifications à partir desquelles le test a été élaboré mettaient l'accent sur les domaines où il fallait avoir un niveau de compétence minimal en anglais pour vivre en société. Le nombre total d'étudiants qui ont participé aux essais sur le terrain aux États-Unis des formulaires B, C et D étaient respectivement de 987 pour le volet entrevue orale et de 632 pour le volet degré d'alphabétisation. Le choix des items à inclure dans le formulaire qui a été retenu pour le test s'est fait essentiellement à partir des résultats statistiques de l'analyse des différents items. Chaque item a été examiné au regard du niveau de difficulté et de la corrélation bisériale (coefficients r). L'analyse a révélé que la

plupart des items des formulaires mis à l'essai pouvaient être conservés. La fiabilité estimative au regard de la cohérence interne des trois formulaires allait de 0,77 à 0,83 pour ce qui est des sous-scores et de 0,96 à 0,97 pour les scores totaux. Ces scores sont très élevés compte tenu du fait que le test est composé en grande partie de questions ouvertes. L'analyse de la fiabilité inter-évaluateurs, qui compare les scores que les différents évaluateurs donnent à un même sujet, a révélé un degré d'uniformité élevé entre les notes des évaluateurs.

La validité apparente du test BEST était aussi élevée puisque son contenu reflète assez fidèlement l'accomplissement en situation réelle des tâches langagières qu'il vise à mesurer. L'analyse de la corrélation entre les scores obtenus et ceux qui avaient été attribués au groupe au départ pour la performance langagière globale des sujets permet de conclure à la validité conceptuelle du test. Pour ce qui est des corrélations entre les sous-échelles, chaque sous-échelle du volet entrevue valide du test témoigne d'une importante corrélation positive avec le score total pour ce volet. De même, il y a une étroite corrélation entre les scores de lecture et d'écriture pour les trois formulaires.

CANADIAN ADULT ACHIEVEMENT TEST

(test canadien de rendement pour adultes)

**The Psychological Corporation,
Harcourt Brace and Company,
555, av. Homer,
Toronto (ON) M8Z 4X6,
Tél.: 1-800-387-7278**

Contenu

Le test canadien de rendement pour adultes (CAAT) est destiné à mesurer le niveau scolaire atteint par un sujet adulte. Il s'agit d'une batterie de sous-tests en vocabulaire, en compréhension de lecture, en orthographe, en calcul arithmétique, en résolution de problèmes, en raisonnement mécanique, en langue, en sciences et en techniques d'étude. Le CAAT comporte quatre niveaux: le niveau A est destiné aux adultes qui ont de un à trois ans d'instruction, le niveau B, aux adultes qui ont de quatre à six ans d'instruction, le niveau C, aux adultes qui ont de sept à dix ans d'instruction, mais qui n'ont pas nécessairement terminé leurs études secondaires, et le niveau D, aux adultes qui ont au moins onze ou douze ans d'instruction et qui ont probablement terminé leurs études secondaires. Les différents sous-tests ne se retrouvent pas tous aux quatre niveaux.

Le CAAT évalue et décrit les compétences génériques de l'apprenant dans l'accomplissement de différentes tâches: décoder des mots, lire des textes de nature scolaire, repérer les fautes d'orthographe, faire des déductions à partir de différents types de textes, faire des additions et des soustractions et savoir résoudre des problèmes liés à la consommation. Les paragraphes ci-dessous donnent au lecteur une idée du contenu de certains des sous-tests.

Le sous-test de vocabulaire est conçu pour évaluer la connaissance et la compréhension de mots que les adultes rencontrent souvent au travail ou dans la vie quotidienne. Le sous-test comprend 34 questions à choix multiples concernant des mots tirés du vocabulaire appliqué ou général, du vocabulaire des sciences physiques et naturelles et du vocabulaire des sciences sociales. Le sous-test de résolution de problèmes comporte 21 problèmes et mesure la capacité du sujet à trouver une solution, à consigner et à récupérer de l'information et à mesurer à l'aide de notions de géométrie. L'apprenant est appelé à résoudre des problèmes comportant des entiers, des décimales, des pourcentages et des fractions du genre de ceux qui se présentent fréquemment dans la vie d'un adulte. Le sous-test de langue, qui est tiré du Stanford Test of Academic Skills (test Stanford de compétences scolaires), est conçu pour mesurer la connaissance fonctionnelle et l'utilisation efficace de la langue anglaise. Le sous-test comprend 54 questions, réparties en quatre sections: recherche documentaire, sensibilité langagière, conventions et disposition en paragraphes

Scores

Le test CAAT permet d'obtenir deux types de scores: des scores liés au contenu et des scores normatifs. La première catégorie comprend notamment score brut objectif. Les scores bruts moyens pour tous les objectifs de chaque niveau du CAAT ont été déterminés pour l'échantillon à qui chaque niveau du test a été administré dans le cadre du programme de recherche sur le CAAT. La performance individuelle ou collective relativement à chacun des objectifs du CAAT peut être évaluée par rapport à ces scores moyens afin de déterminer les points forts et les faiblesses du ou des sujets.

Les scores normatifs servent à comparer la performance du sujet à celle d'un groupe de référence comparable. Les équivalents scolaires et les scores gradués vont des scores normatifs du genre de ceux qui ont déjà été utilisés pour mesurer la performance au CAAT. L'équivalent scolaire est déterminé en comparant le score du sujet à la performance typique d'un certain nombre de groupes d'adultes à qui le test a été administré et qui se trouvaient à un niveau scolaire donné à un mois donné de l'année scolaire. Les équivalents scolaires du CAAT vont du jardin d'enfants à 12,9 années d'instruction, après quoi on passe au niveau postsecondaire. Malgré les limites qu'ils présentent, les scores fondés sur l'équivalent scolaire peuvent donner une indication du niveau approximatif auquel il faudrait débiter l'enseignement.

Les scores gradués du CAAT représentent la performance aux sous-tests et les totaux pour les quatre niveaux. Ainsi, étant donné que le sous-test de compréhension de lecture se retrouve aux quatre niveaux du CAAT, les scores gradués pour ce sous-test vont comparables, qu'il s'agisse du niveau A, B, C ou D. À cause du caractère continu de l'échelle et du fait qu'elle est divisée en intervalles égaux, les scores gradués conviennent tout particulièrement à l'analyse des progrès réalisés pendant la période d'enseignement, peu importe le niveau du CAAT qui a été administré avant et après l'enseignement.

Processus d'élaboration

Parce que le CAAT vise à mesurer le rendement scolaire d'adultes, on a jugé essentiel que l'étalonnage se fasse, non pas avec des enfants d'école, mais avec des adultes. Cependant, malgré tous les travaux consacrés ces derniers temps à la mesure du rendement scolaire des adultes, il n'existe toujours pas de critères acceptables pour

définir la population d'adultes canadiens à qui le CAAT conviendrait. Étant donné l'absence de ces critères, la recherche a porté sur un certain nombre de groupes d'adultes dont les caractéristiques serviraient à définir les utilisateurs de cet instrument. L'échantillon a été établi à partir de listes d'adultes inscrits à des programmes scolaires et professionnels dans divers établissements communautaires et scolaires et à partir de listes d'envoi d'établissements pénitentiaires provinciaux et régionaux. La recherche a permis d'obtenir des informations quant au bien-fondé des questions et des objectifs, quant à la difficulté des questions et à la façon dont elles tenaient compte de la réalité adulte et quant à l'efficacité des options données pour toutes les questions à choix multiples.

Fiabilité. La fiabilité du test est fonction de la mesure dans laquelle il donne systématiquement les mêmes résultats. La fiabilité des scores du CAAT a été déterminée à l'aide de la formule Kuder-Richardson 20. Les estimations de fiabilité vont de 0,84 à 0,95 pour les niveaux A, B et C.

Validité. Pour évaluer la validité du contenu du texte, il faut déterminer s'il mesure les compétences fondamentales dont l'adulte a besoin pour fonctionner dans notre société moderne. Étant donné que les objectifs ont été rédigés de manière à refléter le contenu des programmes d'éducation des adultes à l'échelle du Canada, le CAAT est un instrument valable pour mesurer ces objectifs communs. Il appartient toutefois aux responsables des différents programmes de déterminer la validité du CAAT pour mesurer leurs objectifs. Par ailleurs, l'analyse corrélationnelle a révélé que les tests du CAAT témoignent d'une corrélation positive les uns avec les autres.

MEASUREMENT, READING AND ARITHMETIC

(mesure, lecture et calcul)

**Ramsey Corporation,
Boyce Station Offices,
1050, ch. Boyce,
Pittsburg (PA) 15241-3907,
Tél.: 412-257-0732**

Contenu

Ce test critériel porte sur trois compétences fondamentales, la mesure, la lecture et le calcul. La mesure d'entiers, de moitiés, de quarts, de huitièmes et de sixièmes occupe la première partie du test destinée à des stagiaires qui doivent savoir mesurer avec exactitude. Cette partie du test se compose de 20 items à choix multiples. La partie lecture du test comprend cinq paragraphes exigeant un niveau de lecture de plus en plus élevé. Chacun des cinq passages est suivi de huit questions à choix multiples. La compréhensibilité estimée des passages en question va de 8,0 dans le cas des instructions à 12,0 dans le cas du dernier passage. Le niveau moyen de compréhensibilité est de 10,6.

Les passages liés au travail sont destinés à mesurer la capacité à lire des informations liées au travail du genre de celles qu'on trouve dans divers domaines industriels: fluides hydrauliques, outils d'établi, hydraulique, lubrification et solénoïdes. La partie calcul du test est conçue pour évaluer la capacité à réaliser des opérations arithmétiques simples, additions, soustractions, multiplications et divisions, avec des entiers, des décimales et des fractions. Elle comprend aussi des tableaux et des formules simples. Le test se compose de 40 questions à choix multiples. Chaque partie du test est d'une durée limitée. Pour la mesure, la durée est de 15 minutes; pour la lecture, elle est de 40 minutes; et pour le calcul, elle est de 30 minutes.

Scores

Il s'agit d'un test critériel où un seuil ou un critère de réussite absolu est établi et le score du sujet est déterminé par rapport à ce seuil. Ainsi, un score de 90 p. 100 ou plus correspondrait à un A, un score de 80 à 89 p. 100 correspondrait à un B et ainsi de suite. Des équivalents centiles ont été établis pour les trois parties du texte à partir des scores obtenus par un échantillon de 1 441 métallurgistes.

Processus d'élaboration

Des études de fiabilité ont été réalisées sur d'autres tests de la Ramsey Corporation, comme le Technician Electrical Test, le Technician Mechanical Test, le Building Maintenance Test, l'Electronics Test et l'Electrical Maintenance Trainee Test, mais aucune n'a été faite sur ce nouvel instrument. Sept étapes ont toutefois été prévues pour assurer la validation fondamentale du test. Premièrement, le client fournit une description du travail qui se fait dans son entreprise, laquelle description est reprise suivant les rubriques retenues par la Ramsey Corporation, énoncé des fonctions, travail effectué et conséquences des erreurs. Dix spécialistes du domaine complète la description pour faire en sorte qu'elle soit à jour et qu'elle reflète bien les fonctions. Puis, les dix spécialistes évaluent le travail en question au regard de l'importance des différentes fonctions et du temps en pourcentage qui y est consacré, et leurs évaluations sont ensuite confrontées. La troisième étape consiste à évaluer les niveaux de compréhensibilité, les compétences mathématiques et les autres exigences du travail et du programme de formation. L'entreprise chargée de dresser le test évalue ensuite l'information obtenue de l'analyse du travail et détermine les connaissances, les compétences et les capacités requises de même que les méthodes de mesure. La cinquième étape comprend la révision,

la mise en pages et l'impression du test. Le test est ensuite soumis à un groupe d'au moins quatre spécialistes du domaine pour déterminer s'il reflète bien les connaissances, les compétences et les capacités visées. Puis, on y applique la procédure Angoff pour établir les notes de passage et assurer la révision définitive. La révision est faite par un groupe de cinq à dix spécialistes du domaine qui répondent eux-mêmes aux différents items du questionnaire et qui comparent ensuite leurs réponses avec la bonne réponse pour déterminer quel est le pourcentage de candidate qualifiés qui auraient la bonne réponse. Une fois les indices de difficulté et de la corrélation bisériale de point calculés, la dernière étape consiste à rédiger un rapport sur la validation du contenu suivant la présentation prévue dans l'Uniform Guideline on Employee Selection Procedures (lignes directrices uniformes sur les procédures de sélection des employés). L'entreprise chargée de dresser le test examine le projet de rapport et envoie au client le rapport définitif décrivant dans le détail le processus de validation suivi.

WIDE RANGE ACHIEVEMENT TEST 3

(test de rendement général 3)

Jastak Associates,
C.P. 3410,
Wilmington (DE), 19804-0250,
Tél: 1-800-221-9728

Contenu

Le test de rendement général 3 (WRAT3) est destiné à mesurer les codes nécessaires pour acquérir les compétences fondamentales en lecture, en orthographe et en arithmétique. Cette version récente du test peut être administrée à des adultes de 17 à 75 ans. Chacun des deux formulaires pouvant être utilisés comprend trois sous-tests. Le sous-test de lecture se compose de deux sections, une où le sujet est appelé à lire des lettres et l'autre où il doit lire des mots. Dans la section portant sur les lettres, le sujet est appelé à nommer 15 lettres de l'alphabet, tandis que dans la section portant sur les mots, il est appelé à prononcer 42 mots. Les deux sections du sous-test de lecture doivent être administrées individuellement. Dans le cas des adultes, la section portant sur les mots est administrée en premier. Il s'agit, non pas d'un test de prononciation ou de diction, mais bien d'un test de reconnaissance des mots. Quand les circonstances exigent un test de lecture plus complet, les deux sections du test peuvent être administrées. Un point est accordé pour chaque lettre et chaque mot lu correctement. Les scores bruts sont inscrits sur le questionnaire dans l'espace prévu à cet effet.

Le sous-test d'orthographe comprend des lettres de l'alphabet et des mots qui sont dictés aux sujets. Les deux sections du test se composent de 55 items en tout: 15 lettres de l'alphabet et 40 mots. Quand il s'agit d'adultes, la section portant sur l'orthographe des mots est administrée en premier. Les mots que la personne chargée d'administrer le test doit dicter sont écrits sur des cartes de plastique avec des indications sur la façon de les

prononcer. Un point est accordé pour chaque lettre et chaque mot bien écrit. Les scores bruts d'orthographe sont ensuite inscrits sur le questionnaire dans l'espace prévu à cet effet.

Pour le sous-test d'arithmétique, le sujet doit compter, lire des symboles numériques, résoudre des problèmes qui lui sont exposés oralement et faire des calculs par écrit. Le sous-test se divise lui aussi en deux sections, une section orale comportant 15 items et une section écrite comportant 40 items. Les adultes commencent par la section écrite. Chaque élément compte pour un point et le score brut du sous-test d'arithmétique est inscrit sur le questionnaire dans l'espace prévu à cet effet. Chacun des sous-tests du VVRAT3 est d'une durée de 15 à 30 minutes.

Scores

Plusieurs types de score sont utilisés pour consigner les résultats obtenus au WRAT3. Quatre types de scores sont décrits ci-dessous. Les scores bruts correspondent au nombre total de bonnes réponses. Ils ne peuvent pas donner une mesure précise d'une performance par rapport à une autre. Les scores absolus, qu'il s'agisse d'un seul sujet ou d'un groupe de sujets, mesurent le degré de réussite pour chaque variable quelle qu'elle soit et sans égard au niveau scolaire ou à l'âge. Ces scores conviennent surtout pour les tests administrés avant et après un programme d'enseignement. Les scores normatifs sont ceux qui sont utilisés pour comparer les sujets à eux-mêmes et à d'autres. Il est ainsi possible de déterminer non seulement l'écart entre deux scores, mais aussi l'importance de cet écart. Voici des exemples d'évaluations faites à partir de scores normatifs: moyen supérieur (110-119), moyen (90-109), moyen inférieur (80-89). Les scores correspondent

au niveau scolaire donnent une évaluation un peu plus précise que les scores brute. Dans le cas du WRAT3, ces scores ne permettent de se faire qu'une idée approximative du niveau scolaire atteint en raison de la variation considérable qu'on constate à l'intérieur d'un niveau scolaire, tout comme entre les différents niveaux scolaires. Ces scores devraient être utilisés avec l'information obtenue à partir des scores normatifs ou des scores absolus pour décrire les résultats du test.

Processus d'élaboration

Le WRAT3 est un test normatif fondé sur l'âge qui mesure les capacités de codage scolaire suivantes: reconnaissance de mots, orthographe et calcul arithmétique. Une nouvelle normalisation de cette version du test a été entreprise au début des années 90 à la suite de l'analyse Rasch du degré de difficulté des éléments du test. Cette analyse avait révélé de nombreux cas de double emploi dans les items des deux formulaires du WRAT-R. Le WRAT3 a été administré à un échantillon étalonné aux États-Unis. L'échantillon adulte a été sélectionné à partir d'un modèle d'échantillonnage stratifié avec ajustement en fonction de l'âge, de la région du domicile, du sexe, de l'origine ethnique et du niveau socio-économique. Au total, 4 834 personnes ont été échantillonnées pour tous les groupes d'âge (5-75).

Fiabilité. Quand on mesure une compétence, on devrait normalement pouvoir s'attendre à ce qu'une autre mesure de la même compétence donne des résultats semblables. La fiabilité du test est fonction de la mesure dans laquelle il permet d'évaluer une compétence en particulier de façon systématique. Il existe de nombreuses mesures de la fiabilité d'un test. Voici les résultats des trois mesures qui ont été appliquées au WRAT.

Les coefficients alphas médians des différents formulaires du WRAT vont de 0,85 à 0,95, ce qui est considéré comme souhaitable. Les coefficients de corrélation des scores bruts des trots soul-tests vont aussi très élevés: 0,98, 0,98 et 0,98 respectivement. Selon la méthode test retest, les coefficients de stabilité corrigés vont de 0,91 à 0,98.

Validité du contenu. Les statistiques Rasch permettant d'analyser chaque item ont permis de déterminer dans quelle mesure les items choisis étaient représentatifs des trois domaines de compétence: décodage en lecture, encodage en lecture et calcul arithmétique. Pour chacun des soul-tests, on a constaté que le score obtenu à cet égard était le plus élevé possible, soit 1,00, ce qui tend fortement à confirmer la validité du contenu à l'égard de chacune des trois variables.

Validité conceptuelle. Les données concernant la nature évolutive du WRAT3 révèlent un accroissement constant des scores moyens au fur et à mesure qu'on passe d'un groupe d'âge étalonné à un autre, jusqu'à ce qu'on atteigne les groupes ayant entre 45 et 54 ans. À partir de là, les scores commencent à tomber. Les données relatives à l'ensemble de l'échantillon étalonné permettent de conclure à l'existence d'une corrélation significative entre le score brut au test WRAT3 et l'âge en années et en mois du sujet. L'analyse des corrélations indique par ailleurs que les formulaires du WRAT3 ont une corrélation positive les uns avec les autres. Ainsi, la corrélation entre le test de lecture et le test d'orthographe va de 0,81 à 0,91; la corrélation entre le test de lecture et le test d'arithmétique va de 0,54 à 0,78 et la corrélation entre le test d'orthographe et le test d'arithmétique va de 0,58 à 0,82. Ces corrélations indiquent un rapport allant de modérément à fortement élevé entre les variables.

METRIC SKILLS ASSESSMENT

(évaluation des compétences métriques)

**Lakeshore Technical College,
Gary Heeler,
1290, North Avenue,
Cleveland (WI) 53015,
Tél.: 414-458-4183**

Contenu

L'évaluation des compétences métriques est un test axé sur les compétences qui est conçu pour mesurer les résultats d'une série préétablie d'expériences d'apprentissage métrique. Le test, qui comprend 35 items, est divisé en deux sections. La première section porte sur l'origine, les unités fondamentales de mesure, les définitions et les conversions numériques du système métrique. La deuxième section comprend une série de questions qui amènent le sujet à consigner les mesures obtenues à l'aide de différents instruments de mesure utilisés au travail, comme le micromètre et le pied à coulisse. Le test peut être administré à un ou plusieurs apprenants adultes, et il est d'une durée de 20 minutes environ. Le test peut par ailleurs être utilisé pour l'évaluation préalable ou postérieure à la formation.

Scores

Le test Metric Skills Assessment est un outil critique, de sorte que les progrès accomplis par l'apprenant sont interprétés en fonction du nombre d'unités d'instruction qu'il termine avec succès au niveau critique prescrit. Le test peut être administré préalablement à un programme de formation et, selon le critère utilisé par l'instructeur, il peut déterminer la composante du programme d'études qui convient à l'apprenant. Quand

le test est utilisé à la fin d'un programme de formation, le critère préétabli se révèle généralement correct à 90 % ou 80%.

Processus d'élaboration

La première étape dans l'élaboration du test a été de déterminer avec l'aide d'un groupe de spécialistes des mathématiques les principales compétences qui devaient être acquises. Puis, on a repéré divers matériels pédagogiques offerts sur le marché qui présentaient une multitude d'exercices d'apprentissage permettant d'acquérir chacune des compétences visées. Autrement dit, on a d'abord défini les compétences à acquérir, puis on a conçu le programme d'études de façon à aider l'apprenant à acquérir ces compétences. Le programme d'études comprend six domaines clés. Le premier module consiste en une initiation aux unités de mesures métriques qui sont utilisées dans la vie quotidienne de même qu'une révision des opérations décimales. Ce module comprend les mesures traditionnelles et métriques de même que l'utilisation d'échelles et de galons métriques. Le deuxième module porte sur la conversion entre les unités de mesure métriques et britanniques, tandis que le troisième comprend le filetage métrique et l'identification des appareils de mesure métrique. Dans le quatrième module, on apprend à identifier et à utiliser les instruments de précision, tandis que dans le cinquième, on apprend à utiliser le pied à coulisse. Le programme se termine par une révision du système de mesure métrique et par une réévaluation des compétences.