

Mesure et évaluation en éducation



Par Ridwan Mohamed **OSMAN**



African Virtual university
Université Virtuelle Africaine
Universidade Virtual Africana



NOTE

Ce document est publié sous une licence *Creative Commons*.
http://en.wikipedia.org/wiki/Creative_Commons

Attribution

<http://creativecommons.org/licenses/by/2.5/>

License (abréviation « cc-by »), Version 2.5.



TABLE DES MATIÈRES

I.	Mesure et évaluation en éducation _____	3
II.	Prérequis/ connaissances préalables nécessaires _____	3
III.	Volume horaire/temps _____	3
IV.	Matériel didactique _____	3
V.	Justification/Importance du Module _____	3
VI.	Contenu _____	4
	6.1 Contours _____	4
	6.2 Grandes lignes _____	4
	6.3 Représentation graphique _____	5
VII.	Objectifs Généraux _____	6
VIII.	Objectifs spécifiques aux activités d'apprentissage _____	7
IX	Évaluation des connaissances _____	10
X.	Concepts-clé (glossaire) _____	15
XI.	Lectures obligatoires _____	18
XII.	Activités d'apprentissage _____	25
XIII.	Évaluation sommative _____	90
XIV.	Synthèse du module _____	92
XV.	Auteur du module _____	93



I. Mesure et évaluation en éducation

Par Ridwan Mohamed Osman

II. Prérequis/ connaissances préalables nécessaires

Les cours préalables à ce cours sont psychologie générale, psychopédagogie et psychologie de l'apprentissage

III. Volume horaire/temps

La durée nécessaire à la complétion de ce module est de 120 heures. Ces heures sont réparties de la manière suivante :

- Unité I : 25 heures
- Unité II : 25 heures
- Unité III:25 heures
- Unité IV : 25 heures
- Unité VI : 20 heures

IV. Matériel didactique

Un ordinateur doté d'une connexion Internet, d'un système d'exploitation et d'un logiciel d'application de base.

V. Justification/Importance du module

L'évaluation fait partie intégrante de l'éducation. Elle a un effet sur la manière dont les élèves étudient, sur leur motivation, leur rendement, ainsi que leurs aspirations. Dans un cadre scolaire, elle permet aux inspecteurs et aux fonctionnaires du ministère de déterminer la qualité de l'éducation offerte dans une école donnée, dans une région déterminée ou dans l'ensemble du pays. Dans le cadre d'un programme, elle nous donne une idée de la mesure dans laquelle le module que nous avons mis sur pied – nos moyens – nous permet d'atteindre nos buts et nos objectifs – nos fins. Pour vous qui êtes enseignants, ce module vous permettra de constater l'importance de l'évaluation en éducation. Il vous aidera aussi à décider du meilleur moyen d'évaluer les élèves.



VI. Contenu

6.1 Contours

Ce module – Mesure et évaluation en éducation – a pour but de vous permettre d'être capable de comprendre les principes et les concepts des différents types d'évaluations pédagogiques. Il s'agit d'un cours d'introduction, ce qui signifie qu'il n'est pas attendu de vous que vous deveniez des experts de l'évaluation et des tests en pédagogie. Vous serez par contre aptes à mieux comprendre le processus d'évaluation, de mesure et de tests et vous serez aussi capables d'en mener à bien la planification et l'administration en fonction de principes établis et de pratiques efficaces. Lorsque vous aurez terminé le module, vous serez capables de comprendre ce qui relie et différencie les buts, les objectifs, la mise à l'épreuve et l'évaluation des programmes. Vous aurez aussi une nouvelle perspective sur l'utilisation des résultats aux évaluations.

6.2 Grandes lignes

Unité I : Évaluation en pédagogie

- i. La nature de l'évaluation
- ii. Les types d'évaluation
- iii. Les étapes de l'évaluation

Unité II : Spécification des objectifs

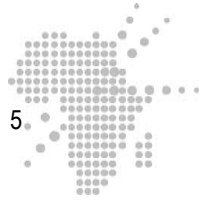
- i. La nature des objectifs
- ii. Classification et énoncé des objectifs

Unité III : Classification et sélection des tests

- i. Classification des tests
- ii. Sélection des tests

Unité IV : Élaboration d'un instrument

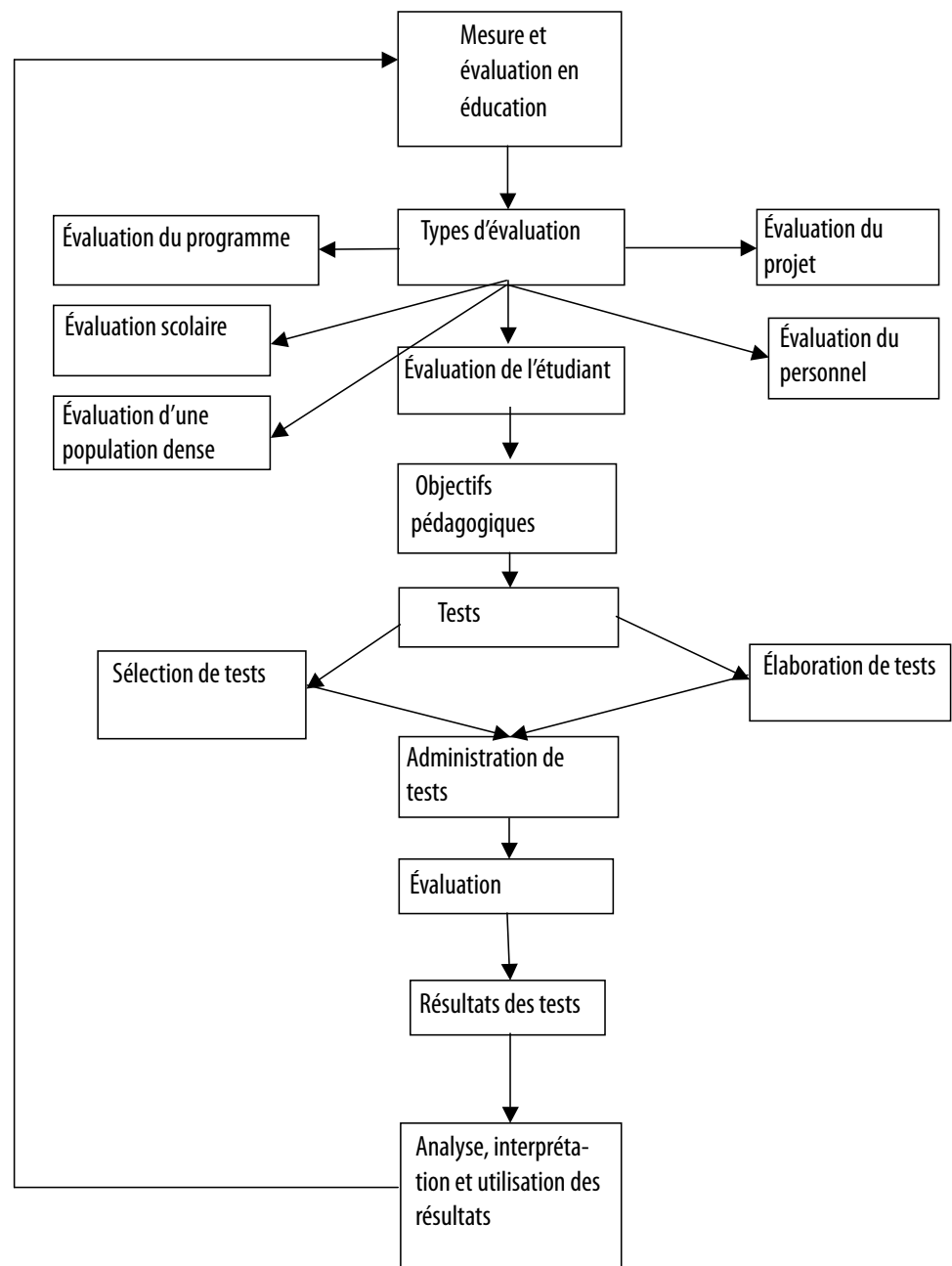
- i. Conception du test
- ii. Construction des items du test
- iii. Conception et reproduction du test
- iv. Validation du test

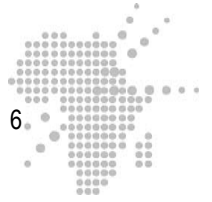


Unité VI : Analyse, interprétation et utilisation des résultats

- i. Analyse
- ii. Interprétation
- iii. Utilisation et exposé des résultats

6.3 Représentation graphique





VII. Objectifs généraux

À la fin de ce cours, l'apprenant devrait :

- (1) démontrer sa compréhension des principes de base de l'évaluation et des tests en pédagogie;
- (2) construire, analyser et améliorer les processus d'évaluation qui jugent de façon adéquate les objectifs d'enseignement généraux ou spécifiques;
- (3) faire la critique et l'interprétation de résultats de tests standardisés;
- (4) intégrer les résultats de l'évaluation à l'évaluation du rendement des élèves;
- (5) expliquer les résultats des évaluations aux élèves, aux parents et aux collectivités;
- (6) construire et interpréter les tests.



VIII. Objectifs spécifiques aux activités d'apprentissage

Unité	Thématiques des unités	Objectifs des activités d'apprentissage
1.	Évaluation en pédagogie	
	La nature de l'évaluation Les types d'évaluation Les étapes de l'évaluation	<ul style="list-style-type: none"> • Définir et décrire l'évaluation • Décrire les étapes principales du processus d'évaluation • Définir et expliquer les termes « mesure », « évaluation » et « test » • Nommer les éléments qu'ont en commun tous les modèles d'évaluation valides • Décrire les différents types d'évaluation en identifiant les principales variables évaluées pour chaque type et pour chaque décision prise • Définir les différentes étapes de l'évaluation et énumérer les activités de chacune de ces étapes
2.	Spécification des objectifs	
	La nature des objectifs Classification et énoncé des objectifs	<ul style="list-style-type: none"> • Nommer les composantes à la base de tout modèle valide d'évaluation des besoins • Nommer trois facteurs à considérer au moment d'établir la priorité des besoins • Nommer et décrire les différents types d'objectifs • Nommer les principes directeurs de la détermination du degré de spécificité requis pour un objectif • Choisir un domaine d'intérêt et rédiger des objectifs connexes pour chacune des catégories principales de la composante cognitive <p>Choisir un domaine d'intérêt et rédiger des objectifs connexes pour chacune des catégories principales de la composante affective</p> <p>Choisir un domaine d'intérêt et rédiger des objectifs connexes pour chacune des catégories principales de la composante psychomotrice</p>



3.	Classification et sélection des tests	
	Classification des tests Sélection des tests	<ul style="list-style-type: none"> • Nommer trois moyens importants de collecter les données et identifier une situation qui s'applique à chacun • Nommer quatre différences importantes entre un test standardisé et un test conçu localement • Nommer les principales catégories de test et décrire leur utilité • Énumérer les caractéristiques d'un bon test • Énumérer les facteurs dont il faut tenir compte au moment de choisir un test parmi d'autres
4.	Élaboration d'un instrument	
	Conception du test Construction des items du test Conception et reproduction du test Validation du test	<ul style="list-style-type: none"> • Identifier et décrire les composantes principales d'un tableau de spécifications • Identifier et décrire les directives générales de construction des items du test • Décrire les types de résultats pour lesquels le compte rendu est le mode d'évaluation à privilégier. • Décrire brièvement deux approches pour noter le compte rendu • Donner deux directives importantes pour la construction de tests objectifs • Nommer un objectif spécifique et un item qui s'y rapporte pour chacun des types de tests suivants : compte rendu, court développement, choix multiple, vrai ou faux et appariement • Décrire les principales approches pour organiser les items du test • Énumérer les principales directives concernant le format des tests • Nommer et décrire les principaux aspects de la validation des tests • Identifier et expliquer les principaux aspects de l'analyse d'items qui s'appliquent à un test normatif • Expliquer les résultats de l'analyse d'items et leur interprétation dans le cadre d'un test normatif • Identifier et expliquer les principaux aspects de l'analyse d'items qui s'appliquent aux tests critériels • Expliquer les résultats de l'analyse d'items et leur interprétation dans le cadre d'un test critériel



5.	Analyse, interprétation et utilisation des résultats	
	<p>Analyse</p> <p>Interprétation</p> <p>Utilisation et exposé des résultats</p>	<ul style="list-style-type: none"> • Décrire les quatre échelons de mesure et donner trois exemples pour chacun • Définir et décrire les statistiques suivantes ainsi que leur interprétation : tendance centrale, variation, loi normale, rangs centiles, écart réduit, note T et corrélation <p>Définir et décrire les équivalences de niveau et d'âge</p> <ul style="list-style-type: none"> • Décrire et expliquer les différents modes d'interprétation des données d'un examen. • Expliquer brièvement le principal usage général des résultats aux tests • Donner quatre principes généraux de rapport • Identifier et décrire brièvement les composantes principales d'un rapport d'évaluation formel



IX. Évaluation des connaissances

Justification

Cette évaluation des connaissances pour le module a pour but de vérifier vos connaissances du contenu de ce module. Après cette évaluation des connaissances, vous serez en mesure de connaître la mesure dans laquelle vous possédez les connaissances et les compétences liées à ce module. Cette évaluation des connaissances vous permettra de juger des efforts que vous devrez fournir dans ce module.

Évaluation des connaissances

1. La collecte et l'analyse systématique de données à des fins décisionnelles est:
 - a. L'évaluation
 - b. La mesure
 - c. L'appréciation
 - d. Le test
2. Le procédé par lequel on évalue jusqu'à quel point une personne ou un objet possède un trait est :
 - a. L'évaluation
 - b. La mesure
 - c. L'appréciation
 - d. Le test
3. L'analyse d'une situation fait partie de quelle étape du processus d'évaluation?
 - a. La planification
 - b. Le processus
 - c. Le produit
4. Les objectifs de maîtrise ont pour but de transmettre :
 - a. La priorité relative des besoins
 - b. Les comportements jugés requis pour réussir
 - c. Les attitudes que devrait avoir tout apprenant
 - d. La complexité des objectifs



5. Parmi les habiletés suivantes, laquelle nécessite des aptitudes physiques?
 - a. Les habiletés cognitives
 - b. Les habiletés affectives
 - c. Les habiletés psychomotrices
6. Les objectifs les plus distincts et mesurables sont :
 - a. La vision
 - b. Les buts
 - c. Les objectifs intermédiaires
 - d. Les objectifs spécifiques
7. Parmi les éléments suivants, lequel devrait faire partie des objectifs pédagogiques?
 - a. Le comportement
 - b. La condition
 - c. Le critère
 - d. Toutes ces réponses
8. La section d'un test dans laquelle les répondants rédigent la réponse est :
 - a. Le compte rendu
 - b. Le choix multiple
 - c. Le vrai ou faux
 - d. L'appariement
9. Les tests de rendement ont pour but de mesurer :
 - a. Les attitudes
 - b. L'intelligence générale
 - c. Le déroulement ou le produit d'une activité
 - d. La personnalité
10. Le test de connaissances mesure :
 - a. Le potentiel des étudiants
 - b. Les caractéristiques personnelles
 - c. La maîtrise des connaissances et des compétences
 - d. La persévérance et la patience



11. L'interprétation des notes suivant le modèle normatif se fait en fonction :
 - a. Des notes des autres étudiants de même niveau
 - b. De la difficulté de l'examen
 - c. De critères préalablement établis
 - d. De l'acquis des étudiants
12. Parmi les éléments suivants, lequel ne caractérise PAS un bon test?
 - a. La validité
 - b. La fiabilité
 - c. La représentativité
 - d. La difficulté
13. La validité du test, c'est :
 - a. Aucune variation de notes, peu importe le correcteur
 - b. La détection des différences de connaissances des étudiants
 - c. Le fait qu'un test mesure ce qu'il doit mesurer
 - d. La cohérence du test
14. Un test souvent administré donnant des résultats semblables est :
 - a. Valide
 - b. Fiable
 - c. Représentatif
 - d. Discriminatoire
15. Le tableau de spécifications a pour but :
 - a. D'assurer la mesure de tous les éléments à évaluer
 - b. D'aider les apprenants à bien comprendre les instructions
 - c. De promouvoir la fiabilité d'un test
 - d. D'équilibrer les difficultés du test
16. Quel est le terme qui désigne le nombre par lequel le suivi des réussites d'un étudiant est assuré?
 - a. La note brute
 - b. Le pourcentage d'exactitude
 - c. L'équivalent de niveau scolaire
 - d. La note standard



17. Parmi les éléments suivants, quel procédé a une incidence sur la fiabilité d'un test?
 - a. L'administration
 - b. La notation
 - c. La normalisation
 - d. Les instructions
 - e. Aucune de ces réponses

18. La validité liée au contenu est un type de validité selon lequel :
 - a. Il a été démontré que le test prévît efficacement les critères ou les indicateurs d'un concept
 - b. Les items du test sont représentatifs des notions que le test doit couvrir.
 - c. Le lien entre les résultats et le trait théorique prédit est démontré
 - d. Aucune de ces réponses

19. Les données d'évaluation peuvent être collectées par:
 - a. L'administration d'instruments standardisés
 - b. L'administration d'instruments conçus localement
 - c. L'observation
 - d. Toutes ces réponses

20. Parmi les éléments suivants, lequel n'est pas un but des tests?
 - a. Décrire le niveau de développement d'un étudiant selon un domaine d'évaluation
 - b. Identifier les forces et les faiblesses d'un apprenant selon la discipline
 - c. Assurer le suivi de la progression annuelle des connaissances de base d'un apprenant.
 - d. Prédire le succès futur d'une personne



Réponses

- 1 a
- 2 b
- 3 b
- 4 c
- 5 c
- 6 d
- 7 d
- 8 a
- 9 c
- 10 d
- 11 a
- 12 d
- 13 c
- 14 b
- 15 a
- 16 c
- 17 e
- 18 b
- 19 d
- 20 d

Pédagogie pour les apprenants

Comme indiqué plus tôt, le but de cette évaluation des connaissances était de vous aider à appréhender votre maîtrise du contenu de ce module. Si vous avez obtenu plus de dix bonnes réponses, vous avez déjà une bonne maîtrise des concepts et des principes de l'évaluation et des tests en pédagogie. Faire les activités de ce module vous aidera à améliorer vos connaissances et vos aptitudes dans ce domaine. Si vous avez obtenu de cinq à dix bonnes réponses, vous n'aurez pas trop de difficulté à étudier ce module. Vous avez une bonne idée de ce que sont les principaux concepts, mais vous devrez tout de même maîtriser les détails. Si vous avez obtenu moins de cinq bonnes réponses, vous devrez fournir temps et efforts pour maîtriser ce module, ce qui vous permettra d'augmenter vos compétences dans le domaine de l'évaluation et des tests en pédagogie.



X. Concepts-clé (glossaire)

L'évaluation est la collecte et l'analyse systématique de données à des fins décisionnelles

La mesure est le procédé par lequel est évalué jusqu'à quel point une personne ou un objet possède un trait spécifique; par exemple, la qualité, le caractère ou une caractéristique.

Préalables : comportement, connaissance ou habiletés que les étudiants devraient avoir.

Les objectifs sont des énoncés spécifiques de ce qui doit être accompli et de la mesure dans laquelle cela doit être accompli. Ils doivent être mesurables et quantifiables.

L'évaluation des besoins est le procédé par lequel les besoins sont identifiés et leur priorité, déterminée.

Les objectifs axés sur les compétences ont pour but de communiquer les comportements jugés nécessaires au succès futur des étudiants en incitant ces derniers à les adopter durant leurs études.

Les objectifs de maîtrise sont des objectifs que devraient atteindre tous les élèves, peu importe leurs aptitudes ou leurs acquis.

La taxinomie des objectifs d'apprentissage a pour but de classer les objectifs selon leur complexité présumée.

Les habiletés affectives sont liées à des objectifs intangibles dont la promotion et la mesure sont difficiles; par exemple les sentiments, les attitudes, les intérêts et les valeurs.

Les habiletés psychomotrices nécessitent des habiletés physiques, dont la coordination musculaire, les habiletés motrices, la manipulation d'objets ou la coordination neuromusculaire.

Buts : les buts sont des énoncés généraux portant sur des résultats à long terme.

Les objectifs intermédiaires sont plus précis que des objectifs généraux, mais pas aussi détaillés que des objectifs spécifiques.

Les objectifs spécifiques ont trait aux résultats les plus distincts.

Un compte rendu est un test pour lequel le nombre de questions est limité, et pour lequel les répondants doivent rédiger une réponse.

Une épreuve objective est une épreuve pour laquelle toute subjectivité de correction est éliminée, du moins en théorie.



Un test standardisé est un test développé par des spécialistes de la mesure qui maîtrisent la matière.

Un test conçu localement est généralement développé localement par l'enseignant dans un but précis.

L'auto-évaluation est une réponse individuelle orale ou écrite.

L'échelle de notation est un instrument comprenant plusieurs items représentant toute une gamme de catégories entre deux extrêmes.

Les tests de connaissances sont des tests qui mesurent le statut d'un individu en fonction de ses aptitudes dans des domaines de connaissance ou de compétences divers.

Les tests de caractère ou de personnalité sont conçus pour évaluer les caractéristiques des personnes ainsi qu'un certain nombre de qualités.

Les tests d'aptitudes mesurent le potentiel de réactivité d'une personne dans des situations auxquelles il pourrait être confronté.

Les normes de rendement sont le critère auquel les résultats sont comparés afin d'être interprétés.

L'évaluation normative est une évaluation, standardisée ou préparée localement, qui rapporte et interprète les notes en les comparant aux autres notes obtenues pour le même test.

Le test critériel est un test qui rapporte et interprète la note en fonction d'un standard absolu.

L'objectivité du test est le fait que, peu importe le correcteur, le résultat du test reste le même.

La discrimination est la mesure dans laquelle un test détecte ou mesure les petites variations de rendement.

La représentativité est la mesure dans laquelle un test reflète les objectifs principaux du cours.

La validité consiste en le fait qu'un test mesure ce qu'il doit mesurer.

La validité du contenu est la mesure dans laquelle les items d'un test correspondent à l'étendue des items qu'un test devrait normalement couvrir.

La validité critérielle est la mesure dans laquelle un test réussit effectivement à prédire des critères ou à indiquer une construction.

La validité concourante est rencontrée lorsque la mesure des critères a lieu en même temps que l'obtention des notes.



La validité prédictive est rencontrée lorsque la mesure des critères a lieu après le test.

La validité conceptuelle est la mesure dans laquelle un test démontre un lien entre les résultats obtenus au test et le trait théorique prédit.

La fiabilité est la cohérence des mesures. Un test est considéré fiable lorsque les mêmes résultats sont obtenus de manière répétée.

Un tableau de spécifications est essentiellement un projet de test dont le but est de s'assurer que tous les résultats visés, et seulement les résultats visés, sont évalués et que le test contienne un nombre conséquent de questions pour chaque item évalué.

La validation du test est le procédé par lequel un test est administré et revu jusqu'à l'obtention d'un niveau de validité et de fiabilité acceptable.

Une note brute est le nombre de bonnes réponses obtenues par un étudiant (en supposant que chaque question vaut un point).

Le pourcentage d'exactitude est obtenu par la division de la note brute par le nombre total de questions et ce résultat est multiplié par 100.

L'équivalent de niveau scolaire est le nombre par lequel le suivi des réussites d'un étudiant est assuré.

Le rang centile est le pourcentage d'élèves d'un groupe précis qui a obtenu une note inférieure à celle de l'étudiant.



XI. Lectures obligatoires

Lecture #1

Titre : Designing Evaluation for Education Projects

Référence complète : *wateroutreach.uwex.edu/use/documents/NOAAEvalmanual-FINAL.pdf*

Résumé

La lecture couvre les principes fondamentaux de l'évaluation, présente des analyses variées sur une panoplie de sujets : types d'évaluation, façons de collecter les renseignements, utilisation d'un évaluateur à l'extérieur et considérations morales liées à la collecte de données des participants au programme. Ces informations veulent répondre aux questions concernant l'évaluation du projet et fournir des conseils lors de l'utilisation de l'évaluation à titre d'outil pour l'amélioration du projet.

Justification

Beaucoup de temps, d'efforts et de ressources sont consacrés au développement et à la mise en œuvre des projets d'enseignement. Bien évidemment, le but est de concevoir des projets efficaces qui peuvent servir de modèles d'excellence. Peu importe si le projet est une marche d'une heure pour voir les espèces en voie de disparition, un festival du jour de la Famille, l'observation des ressources marines ou un forum communautaire, le but, qui est de fournir des expériences pédagogiques de qualité, ne change pas. Le but de cette lecture est de vous aider – vous les enseignants, les administrateurs, les gestionnaires de projet et les décideurs de demain – à comprendre le processus d'évaluation des projets pour améliorer leur efficacité et leur incidence positive.



Lecture #2

Titre: Bloom's Taxonomy of Educational Objectives

Référence complète: www.humboldt.edu/~tha1/bloomtax.html

Résumé

La lecture présente au lecteur un survol de la taxonomie des objectifs pédagogiques de Bloom. Chacun des trois domaines –cognitif, affectif et psychomoteur – est expliqué et les sous-catégories de chacun sont fournies. De plus, le lecteur aura des listes de verbes qui l'aideront à établir des objectifs adéquats pour chaque catégorie.

Justification

Dans ce module, ce document à lire favorisera la compréhension de l'apprenant en ce qui a trait aux trois domaines de la taxonomie de Bloom. Les verbes fournis aideront l'apprenant à établir des objectifs appropriés pour chaque catégorie.

Lecture # 3

Titre: Taxonomy of Educational Objectives

Référence complète: www.humboldt.edu/~tha1/bloomtax.html

Résumé

Cet article fournit un exemple de chaque taxonomie des objectifs pédagogiques de Bloom à l'apprenant. De façon plus détaillée, les exemples facilitent la compréhension des concepts.

Justification

Les exemples accompagnant la taxonomie vous aideront à établir des objectifs efficaces et à préparer des questions ou des activités afin de mesurer l'atteinte de ces objectifs.



Lecture # 4

Titre: Writing Instructional Objectives

Référence complète: www.nacls.org/docs/announcement/writing-objectives.pdf

Résumé

Au début, l'article souligne les différences des objectifs pédagogiques, des descriptions de cours et des préalables. Ensuite, il fournit un guide facilitant l'établissement des objectifs pédagogiques. Les règles de cet article sont aussi liées aux domaines de Bloom.

Justification

Cet article met l'accent sur les objectifs pédagogiques, qui constituent d'importants points de départ pour l'évaluation de l'élève. Les futurs professeurs et les autres apprenants de ce module seront aptes à établir des objectifs spécifiques quantifiables qui tiennent compte de tous les domaines d'apprentissage.

Lecture # 5

Titre: The Basics of Instructional Objectives

Référence complète: *ed.fullerton.edu/SecEd/Credential_Prog/EDSC_440S_Materials/Instructional_Objectives.ppt*

Résumé

Ce document à lire explique la nature des objectifs pédagogiques et leurs buts en enseignement et en évaluation. De plus, ce document distingue les quatre principaux éléments des objectifs pédagogiques (à savoir le public, le comportement, les conditions et la formation). Aussi, cet article fournit des exemples d'objectifs quantifiables. À la fin du texte, des distinctions sont faites entre les différents types d'objectifs (résultat contre processus, élève contre professeurs, et général contre spécifique).

Justification

Cette lecture présente un aperçu des objectifs pédagogiques et de leurs éléments à l'apprenant. Elle lui permettra d'identifier les éléments des objectifs pédagogiques et de vérifier si ses objectifs qu'il a déterminés se conforment à cette norme.



Lecture # 6

Titre: *The Case Against Standardized Tests* **Completer Reference:** www.al-fekohn.org/books/tcast.htm

Résumé

Cet article critique les tests standardisés en exposant brièvement les principales faiblesses constatées dans les ouvrages sur ce thème. À quel point les notations des tests sont-elles valables pour prédire les notes? Sont-elles valables ou non pour prédire l'accomplissement actuel? Les tests ont-ils un préjugé défavorable envers certains membres de la société? Cet article passe en revue les ouvrages critiques et exhaustifs consacrés au sujet des tests standardisés pour tenter de répondre à ces questions.

Justification

Cet article permettra à l'apprenant de voir le côté négatif des tests standardisés. Les faiblesses exposées dans cet article se manifesteront concrètement en pratique lorsque l'apprenant sera confronté à ce choix difficile : tests standardisés ou tests conçus localement?

Lecture # 7

Titre: *Developing a Table of Specifications*

Référence complète: academicaffairs.cmich.edu/caa/assessment/posters/TableofSpecifications.pdf

Résumé

Dans cet article, les aspects communs du tableau de spécifications sont présentés et expliqués. Une attention particulière est portée à la validité du contenu et du processus. On retrouve également des modèles de tableaux de spécifications dans l'annexe de l'article. Ces tableaux sont représentatifs de plusieurs domaines des programmes éducatifs à l'enseignement secondaire.

Justification

Le tableau de spécifications est un outil important permettant au professeur de mettre au point des examens qui représentent bien le contenu et les objectifs qu'ils sont censés évaluer. Le but de cet article est de faciliter l'élaboration par les apprenants de tableaux de spécifications pour les examens en classe.



Lecture # 8

Titre: Test Construction: Some Practical Ideas

Référence complète: www.utexas.edu/academic/cte/sourcebook/tests.pdf

Résumé

Cet article présente certaines idées pratiques sur l'élaboration de tests et la construction d'items qui évaluent ce qu'ils sont censés évaluer de la meilleure façon possible. En tout, il y a 12 directives qui traitent du processus de la conception de tests. Des concepts comme des plans détaillés de tests, d'autres genres de questions et une analyse des éléments sont abordés plus en profondeur.

Justification

Cet article vise à fournir à l'apprenant un aperçu sur la façon de construire les tests actuels.

Lecture # 9

Titre: Improving Essay Tests

Référence complète: www.theideacenter.org/sites/default/files/Idea_Paper_17.pdf

Résumé

Cet article, qui met l'accent sur l'élaboration de comptes rendus, donne une bonne vue d'ensemble de ce genre d'examen. Il commence par décrire le compte rendu et ensuite, il fournit un panorama des forces et des faiblesses de ces examens. Ce texte émet aussi des recommandations sur le moment approprié d'utiliser ces examens et la façon de l'élaborer. À la fin de l'article, le processus de la notation des comptes rendus est abordé.

Justification

Cet article aidera les apprenants à se faire une idée de la façon dont il faut concevoir des questions à réponse élaborée qui sont adéquates pour des examens en classe. Aussi, le lecteur aura un aperçu des principales faiblesses qui freinent l'efficacité des examens à questions ouvertes, qui sont des outils pédagogiques.



Lecture # 10

Titre: Designing Objective Test Questions: An Introductory Workshop Référence complète: <http://www.caacentre.ac.uk/dldocs/otghdout.pdf>

Résumé

Cet article présente les différents types de questions pour les tests objectifs et la façon de les construire. Les types de questions abordés comprennent les questions à choix multiples (QCM), les « vrai ou faux », la raison de l'affirmation, les questions à réponses multiples (QRM) et les questions d'appariement.

Justification

Ce lien aidera les apprenants à se faire une idée générale des différents types de questions pour les tests objectifs. Dans chaque catégorie, des exemples sont donnés pour vous faciliter la tâche lors de l'élaboration des questions de ce genre.

Lecture #11

Titre: **How to Prepare Better Multiple – Choice Test Items: Guidelines for University Faculty**

Référence complète: <http://testing.byu.edu/info/handbooks/betteritems.pdf>

Résumé

Cette brochure est consacrée entièrement aux questions à choix multiples, qui sont le type de questions d'examen sur les objectifs le plus important. Elle traite de la structure des QCM, des avantages et des faiblesses, des moments appropriés pour les utiliser et des différents types de QCM. L'article explique comment les QCM peuvent être employées pour évaluer les compétences de plus haut niveau. À la fin de cette brochure, on retrouve des conseils pour l'élaboration de ce type de questions d'examen.

Justification

Après avoir étudié exhaustivement cette brochure et mis en pratique ce qui avait à l'intérieur, l'apprenant devrait être capable de distinguer les objectifs pédagogiques qui sont évalués le plus pertinemment par les QCM, d'évaluer les QCM en se servant des critères couramment acceptés et en soulignant les défauts des QCM et de recommander des façons de les améliorer. Par-dessus tout, l'apprenant devra être apte à construire des QCM bien écrites qui évaluent des objectifs spécifiques.



Lecture #12

Titre: Best Practice in Mathematics: Using Test Results to Inform Instruction and Improve Student Achievement

Référence complète: http://www.ctb.com/media/articles/pdfs/resources/ctb_best_practice_in_math.pdf

Résumé

Cet article examine les façons que les professeurs peuvent utiliser les résultats aux examens pour donner des instructions et améliorer le rendement scolaire. Un certain nombre de conseils sont donnés au professeur qui souhaite utiliser les résultats des examens. Il y a aussi un exemple de bulletin.

Justification

Cet article donnera des pistes aux professeurs lorsqu'ils penseront à des façons d'utiliser les résultats des examens; de la sorte, leur pédagogie sera liée aux besoins et aux faiblesses des élèves.

Lecture #13

Titre: Descriptive Statistics

Référence complète:

http://www.sagepub.com/upm-data/9881_040143ch02.pdf

Résumé

Cet article présente les concepts sous-jacents aux statistiques descriptives en donnant des exemples complexes tirés des sciences sociales. On retrouve des types de données, des statistiques descriptives visuelles et des statistiques descriptives numériques.

Justification

À la suite de cette lecture, les apprenants auront une grande compréhension des statistiques descriptives et seront en mesure d'utiliser ces concepts pour leurs examens quotidiens.



XII. Activités d'apprentissage

Activité d'apprentissage # 1

ÉVALUATION PÉDAGOGIQUE

Résumé

La première activité de ce module vise l'acquisition chez l'élève d'une compréhension élémentaire de l'évaluation pédagogique et de l'examen. La terminologie de base relative au sujet sera abordée. De plus, des types d'évaluation devraient être considérés. L'activité d'apprentissage couvrira également les différentes phases de l'évaluation, et d'abord par la préparation et la finalisation du processus. Les lectures et les exercices intégrés à cette activité d'apprentissage devraient permettre à l'apprenant d'atteindre les objectifs établis pour les unités correspondantes du module.

Liste de concepts-clé

L'évaluation est la collecte et l'analyse systématique de données à des fins décisionnelles

La mesure est le procédé par lequel est évalué jusqu'à quel point une personne ou un objet possède un trait spécifique; par exemple, la qualité, le caractère ou une caractéristique.

Préalables : comportement, connaissance ou habiletés que les étudiants devraient avoir.



Contenu

Évaluation et mesure

L'évaluation est une partie intégrante de tous les systèmes d'éducation, et ce, à toutes les étapes. Elle permet aux éducateurs, aux professeurs, aux gestionnaires, aux décideurs et à la communauté de se faire une idée de ce qui manque et ce qui est accessible. L'évaluation peut être définie de deux manières, tout dépendant de ce que nous voulons obtenir à la fin de l'exercice.

Elle est la collecte et l'analyse systématique de données afin de déterminer si les objectifs ont été atteints et dans quelle mesure ils ont été atteints ou sont en voie d'être atteints.

L'évaluation est la collecte et l'analyse systématique de données afin de prendre des décisions.

La première partie de ces définitions (collecte et analyse systématique de données) est commune aux deux définitions formulées ci-dessus. Toutefois, les principales différences entre ces définitions touchent à la question d'inclure ou non les décisions ou les jugements en tant que partie intégrante de l'évaluation. La seconde définition semble être plus inclusive et n'exclut pas les activités impliquées dans la première définition.

Dans la plupart des cas, l'évaluation veut répondre aux questions qui nécessitent un niveau d'évaluation. Des questions comme:

- a) Est-ce que le programme spécial en vaut la peine pour ce qu'il coûte?
- b) Est-ce que le nouveau programme d'études en lecture à l'essai est meilleur que le précédent?
- c) Est-ce que M. Ahmed devrait être placé dans un programme pour les surdoués ?

nécessite la collecte et l'analyse de données et l'interprétation de ces données à l'égard d'un critère ou plus. Plus les critères sont objectifs, plus ils sont efficaces, car il y a inévitablement une part de subjectivité étant donné que les gens prennent la décision finale. Pour obtenir des réponses aux questions précédentes, l'évaluation ne consiste pas uniquement à déterminer si quelque chose est « bon » ou « mauvais ». La fonction ou le but de l'évaluation est de

- déterminer la position actuelle de l'objet de l'évaluation
- comparer le statut avec un ensemble de normes ou de critères
- effectuer un choix afin de prendre une décision



À la fin du processus, il se peut qu'il y ait seulement deux choix ou une combinaison d'activités et de programmes complexes conçus comme intervention pour la situation actuelle. En général, le processus d'évaluation comporte la détermination des types de données qui doivent être collectées, la détermination des personnes, du ou des groupes qui constitueront l'échantillon, la collecte, l'analyse et l'interprétation des données et la prise de décision.

Il est crucial que les décisions soient basées sur une procédure valide, car bien que toutes les décisions ne revêtent pas la même importance, chacune a une incidence directe ou indirecte sur les élèves. Plus une conséquence est importante, plus la décision est importante. Toutefois, chaque décision pédagogique devrait être prise logiquement et objectivement, et devrait se baser sur les données les plus pertinentes, et ce, dans la mesure du possible.

Les données collectées durant le processus d'évaluation sont aussi bonnes que les mesures sur lesquelles les mesures sont fondées. La mesure est le procédé par lequel est évalué jusqu'à quel point une personne ou un objet possède un trait spécifique, par exemple, la qualité, le caractère ou une caractéristique.

Les mesures permettent une description plus objective en ce qui concerne les traits de caractère et facilitent les comparaisons. Ainsi, au lieu de dire que Kimani est maigre pour son âge et sa taille, nous pouvons dire que Kimani a 16 ans, mesure 5 pieds et 8 pouces et pèse seulement 85 livres. Aussi, à la place de dire que Kimani est plus intelligente que Juma, nous pouvons dire que Kimani possède un QI de 125 et que le QI de Juma est de 88. Dans chaque cas, l'utilisation des chiffres est plus précise, plus objective et moins sujette à l'interprétation que les descriptions verbales qui veulent dire la même chose.

Est-ce que la quantification de tous les traits d'intérêts est possible? Chez les professionnels, il y a divergence d'opinions sur cette question. Pouvons-nous vraiment quantifier des qualités indéfinissables comme l'empathie, l'appréciation, la motivation ou les intérêts? Pouvons-nous vraiment quantifier des valeurs et des attitudes? La réponse est « oui », mais c'est difficile. Les défenseurs de cette position utilisent l'argument suivant : si quelque chose existe, il existe en quantité; s'il existe en quantité, il peut être mesuré. Si nous acceptons cette logique, aux yeux des éducateurs, tout trait d'intérêt peut être quantifié. Ainsi, le but de la mesure psychopédagogique est de savoir dans quelle mesure une personne ou une entité possède « quelque chose ».

Admettre que, du moins théoriquement, toutes les choses peuvent être mesurées et mesurer toutes ces choses sont deux processus complètement différents. En éducation, il est difficile d'avoir des mesures valides. Le principal problème est que, à l'exception de la mesure des caractéristiques physiques comme la taille et le poids, toute mesure est indirecte; il n'y a pas de critère ou de barème pour mesurer



des traits comme l'intelligence, la réussite ou l'attitude. L'évaluation de ces traits doit nécessairement se baser sur la déduction. En outre, le problème est aggravé par le manque d'instruments bien validés. Dans certains domaines, le problème est moins sérieux que dans d'autres.

Le terme « mesure » ne rime pas avec l'administration d'examens en classe. Les données peuvent être collectées par l'observation et l'analyse et l'évaluation d'un produit. Dans certains cas, on a peut-être déjà accès aux données nécessitées et celles-ci peuvent se trouver dans des dossiers. Toutefois, à maintes occasions, certaines combinaisons de tests standardisés ou de tests autoconçus, ou des deux, sont requises.

Types d'évaluation

Lorsque nous faisons référence aux types d'évaluation, nous faisons allusion aux différents processus, produits et personnes qui peuvent être évalués, notamment les élèves, les programmes éducatifs, les écoles, les systèmes scolaires, les populations denses, les programmes ou les projets spéciaux et le personnel. Quand nous parlons des différents types d'évaluation, cela ne signifie pas qu'il y a plusieurs processus d'évaluation différents. À la base, le processus d'évaluation reste le même, peu importe ce qui fait l'objet de l'évaluation. Ce qui diffère, c'est ce qui est évalué, comment le processus d'évaluation est appliqué et le type de décisions effectuées. Les différents types de données collectées varieront en fonction de ce qui est évalué ; divers critères s'appliqueront aux différentes données et des types différents de décisions seront effectuées. Mais à la base, le processus d'évaluation reste le même et les mêmes concepts et principes généraux de mesure et d'évaluation sont en vigueur.

a) Évaluation de l'élève

La réussite est l'une des variables qui font l'objet de l'évaluation de l'élève; les autres variables importantes comprennent l'aptitude, l'intelligence, la personnalité, les attitudes et les intérêts. Afin d'évaluer la réussite, des tests standardisés et conçus par le professeur sont donnés; les projets, la procédure et les présentations orales sont évalués; et de l'observation formelle et informelle est effectuée. Un professeur se sert des données liées au rendement non seulement pour évaluer le progrès réalisé par l'élève, mais également pour évaluer sa propre pédagogie. En d'autres mots, le processus d'évaluation des élèves donne de la rétroaction au professeur. La rétroaction sur le progrès actuel de l'élève peut aussi guider le professeur en vue des activités pédagogiques ultérieures. Toutes sortes de décisions effectuées ont une incidence directe sur l'élève et toutes ces décisions ne sont pas prises par le professeur. Celui-ci choisit :



- si l'élève a atteint ses objectifs, ou dans quelle mesure, et fournit un enseignement correctif adéquat le cas échéant;
- si l'élève développe un potentiel;
- les attentes vis-à-vis un élève ou un groupe d'élèves en particulier;
- si l'enfant a des besoins particuliers qui ne peuvent être satisfaits en classe ordinaire et peut recommander qu'un enfant soit placé dans un environnement adapté.

Une multitude d'autres décisions prises ayant trait aux élèves ne relèvent pas du professeur et celui-ci ne peut qu'émettre des recommandations. Ces décisions comprennent :

- la place de l'enfant dans un programme particulier;
- l'admission d'un élève en classe supérieure;
- le placement d'élèves et leur admission dans d'autres établissements scolaires.

b) Évaluation du programme éducatif

L'évaluation du programme éducatif comporte l'évaluation de tout programme ou matériel pédagogique et celle de facteurs comme les stratégies d'enseignement, les manuels, le matériel audiovisuel et la disposition physique et organisationnelle. L'évaluation du programme d'enseignement peut aussi comprendre l'évaluation d'un programme complet ou celle d'un petit aspect d'un programme éducatif en entier, comme pour un film. Bien que les programmes continus fassent l'objet d'une évaluation, l'évaluation du programme d'enseignement est généralement liée à l'innovation, une approche nouvelle ou différente; l'approche peut être générale ou spécifique à un domaine donné.

L'évaluation du programme d'études implique généralement des critères internes et externes, et des comparaisons. L'évaluation interne voit si le nouveau processus ou produit accomplit les objectifs exposés; en d'autres mots, elle vérifie s'il fait ce qu'il prétend faire et évalue aussi les objectifs. L'évaluation externe voit si le processus ou le produit fait quelque chose en mieux que d'autres processus ou produits particuliers.

Outre la réussite de l'élève, il y a un certain nombre d'autres facteurs qui devraient être pris en considération lors de l'évaluation du programme éducatif. Les deux facteurs les plus importants sont l'attitude et le coût. La recherche a prouvé qu'il y a une relation proportionnelle entre l'attitude du professeur envers le programme d'enseignement et son efficacité ultime.



Le principal problème lié à l'évaluation du programme est qu'il est très difficile de comparer équitablement l'efficacité d'un programme ou d'une approche avec d'autres programmes ou approches. Même si deux programmes traitent de la même discipline, ils peuvent posséder des objectifs très différents, et trouver un test ou d'autres mesures qui ont la même justesse ou validité n'est pas une mince tâche. Si un programme d'études doit être comparé à un autre, les objectifs de chacun doivent être examinés attentivement : si on ne peut trouver aucune mesure qui s'applique adéquatement aux deux programmes, alors il faut en concevoir une nouvelle.

c) Évaluation de l'école

L'évaluation d'une école tient compte de l'évaluation de la totalité de son programme d'éducation et comporte la collecte de données sur tous les aspects du fonctionnement de l'école. Le but de l'évaluation de l'école est de déterminer dans quelle mesure les objectifs sont atteints et de cerner les points forts et faibles du programme en entier. Les renseignements recueillis des programmes d'enseignement fournissent de la rétroaction qui guide les activités prochaines de l'école et qui motive des décisions concernant la répartition des ressources.

Un aspect principal de l'évaluation de l'école est le programme de tests pour l'école; plus le programme de test est complet, plus les données recueillies à partir de ce programme sont précieuses. Un programme de tests pour l'école devrait comporter la mesure de la réussite, de l'aptitude, de la personnalité et de l'intérêt. Les tests choisis pour une école doivent concorder avec les objectifs de l'école et être adéquats pour les élèves qui les subiront.

En général, l'évaluation de l'école comprend plus que l'administration de tests aux élèves; elle peut nécessiter des questionnaires, des entrevues et des observations à la suite des données collectées auprès de toutes les personnes au sein de la communauté scolaire, y compris les membres de la direction, les professeurs et les conseillers.

d) Évaluation de populations denses

L'évaluation de populations denses comporte l'évaluation de la position actuelle et le progrès éducatif d'un grand nombre d'élèves, généralement répartis dans une grande région géographique. Les programmes d'évaluation couvrant tout un État sont généralement fondés sur l'idée que le système éducatif de l'État est responsable de la réalisation de certaines compétences fondamentales exigées chez l'élève pour fonctionner efficacement dans notre société, et sur l'idée que les programmes conçus pour favoriser la réalisation de ces compétences devraient



être efficaces et économiques, dans la mesure du possible. Un des principaux buts de l'évaluation de l'État est de diffuser des renseignements aux décideurs de l'État et des régions au sujet du caractère adéquat des programmes éducatifs de base. L'évaluation couvrant tout un État comporte généralement l'évaluation d'un nombre minimum d'objectifs éducatifs et d'objectifs facultatifs qui se servent de tests critériels et normatifs.

e) Évaluation de projets et de programmes particuliers

Les projets et les programmes particuliers comprennent tous les efforts concertés qui ne font pas partie, à vrai dire, des programmes d'éducation ordinaires; généralement, ils sont modernes et la durée de leur existence varie en fonction de leur succès. Que cela soit exigé ou non, la tenue d'une évaluation est d'un intérêt primordial pour un projet ou un programme étant donné qu'elle constitue la seule façon valide de vérifier son efficacité.

f) Évaluation du personnel

L'évaluation du personnel comporte l'évaluation de toute personne responsable, directement ou indirectement, des résultats pédagogiques : professeurs, gestionnaires, conseillers, etc. On a appris que ce secteur d'évaluation est très compliqué; en effet, il est difficile de déterminer les comportements qui doivent être évalués. La meilleure solution au problème de l'évaluation du personnel est de collecter de nombreuses données justes provenant de sources très diversifiées, dans la mesure du possible.

Phases de l'évaluation

L'évaluation est un processus continu et contrairement à ce que la plupart des gens pensent, ce n'est pas ce qui est effectué à la fin qui est évalué. L'évaluation devrait être planifiée avant le déploiement de tout effort et devrait faire partie de l'activité, du début à la fin. Généralement, il y a une série de fins temporaires dans un cycle continu. Par exemple, pour l'évaluation de l'élève, nous commençons par un ensemble d'objectifs pédagogiques. Ensuite, nous évaluons le rendement, ce qui est une fin temporaire dans le cycle pédagogique. En fonction des résultats, nous réévaluons nos objectifs et nos stratégies et nous poursuivons. Ainsi, le processus est cyclique, et la rétroaction obtenue dans un cycle guide le prochain cycle. Outre les résultats, nous évaluons d'autres aspects; chaque étape du processus est soumise à l'évaluation et commence par les objectifs.



Le processus d'évaluation comporte la prise de décision. Tout effort pédagogique entraîne un ensemble de décisions qui doivent être effectuées – des décisions sur les objectifs, les stratégies, les mesures, etc. Ces décisions diverses peuvent être classées en fonction du moment où elles ont été prises, c'est-à-dire l'étape de l'activité à l'étude. Ainsi, chaque étape de l'évaluation comporte différentes sortes de décisions. Logiquement, nous pouvons distinguer trois étapes : l'étape de la planification, du processus et du produit.

L'étape de la planification

Cette première étape de l'évaluation s'effectue avant la mise en œuvre actuelle et nécessite la prise de décisions au sujet du plan d'action qui sera adopté et de ses buts. L'étape de la planification met en jeu un certain nombre de processus qui sont examinés ci-dessous.

a) Analyse de situation

Analyser la situation comme elle l'est présentement constitue la première initiative afin d'établir les paramètres de l'effort. Cette initiative comprend des activités comme la collecte de renseignements généraux et l'évaluation des contraintes actuelles. En conséquence, le professeur peut être appelé à examiner les dossiers commutatifs de ses élèves afin d'avoir un cadre de référence fondé sur leurs habiletés et leurs antécédents. Une fois les paramètres établis, on peut élaborer des buts et des objectifs réalistes.

b) Spécification des objectifs

Les buts sont des énoncés des objectifs, ou les résultats souhaités, et ne sont pas quantifiables directement. Chaque objectif doit être divisé en un objectif spécifique ou plus, et ces objectifs seront mesurables. Ainsi, les objectifs sont des énoncés spécifiques concernant ce qui devrait être accompli et la bonne façon d'y parvenir et sont exprimés en résultats mesurables et quantifiables. Les objectifs peuvent être axés sur le processus ou le produit. Les objectifs de processus décrivent les résultats souhaités durant le déploiement de l'effort et ils sont liés au développement et à réalisation. De leur côté, les objectifs connexes au produit décrivent les résultats voulus à la suite de l'effort.

Les objectifs guident toutes les activités ultérieures et l'atteinte des objectifs est évaluée en dernier lieu. Les objectifs, d'enseignement ou de programme, constituent les bases de toute activité d'évaluation subséquente et par conséquent, il est essentiel qu'ils soient évalués selon la pertinence, le caractère mesurable, le contenu et la précision technique.



c) Spécification des préalables

Les objectifs nécessitent une procédure unique en ce qui concerne l'évaluation de l'élève. Dans la plupart des cas, les caractéristiques d'un ensemble donné d'objectifs pédagogiques sont basées sur l'idée selon laquelle les élèves ont acquis auparavant certaines compétences et connaissances. Si cette idée est fautive, les objectifs sont donc inadéquats. Les comportements anticipés sont appelés les comportements préalables ou d'entrée. Ces préalables doivent être précisés et mesurés, car la pédagogie et l'évaluation systématiques l'exigent. L'évaluation des comportements d'entrée est particulièrement importante au début de toute unité d'instruction. Pour parvenir aux préalables, vous vous posez la question : « Qu'est-ce que tous les élèves doivent savoir ou être en mesure d'accomplir avant l'enseignement afin de profiter de celui-ci et d'atteindre tout objectif? »

d) Choix et élaboration des instruments de mesure

La collecte de données pour déterminer le degré d'atteinte des objectifs exige l'administration d'un instrument ou plus. De tels instruments doivent être conçus ou choisis. Le choix des instruments implique l'étude de ceux qui sont accessibles et le choix du meilleur. Dans ce cas-ci, le meilleur veut dire celui qui est le plus approprié pour vos objectifs et vos utilisateurs. La mise au point d'un bon instrument peut prendre beaucoup de temps, et d'efforts, et exiger des compétences particulières. Pour ce faire, une formation en évaluation est nécessaire.

e) Présentation des stratégies

D'ordinaire, les stratégies sont des approches qui encouragent l'atteinte d'un objectif ou plus. Il peut y avoir des stratégies pour la pédagogie, les programmes éducatifs et les programmes. Chaque stratégie comporte un certain nombre d'activités spécifiques, et généralement, on peut choisir parmi une multitude de stratégies. L'utilisation de ces stratégies doit être planifiée afin de faire en sorte que les ressources nécessaires soient accessibles. Les stratégies qui doivent être le fruit d'une mûre réflexion avant la tenue de l'évaluation comprennent l'analyse des tâches, la révision des concepts, la séquentiation et l'utilisation de la rétroaction et la pratique.

f) Préparation d'un calendrier général

La préparation d'un calendrier général réaliste est importante pour tout type d'évaluation; rares sont les fois où nous avons autant de temps que nous voulons pour mener notre évaluation. En fait, un calendrier général comprend une liste des principales activités de l'effort d'évaluation proposé, du temps correspondant à chaque début prévu et du temps d'exécution pour chaque activité. Vous devriez



prévoir suffisamment de temps; de la sorte, vous pourriez encore respecter votre dernière échéance en cas de contretemps.

Étape du processus

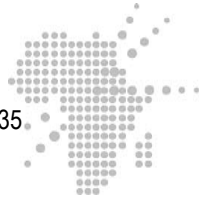
L'étape du processus nécessite des décisions prises en fonction d'événements qui se produisent durant la mise en œuvre actuelle de la pédagogie, du programme ou du projet prévu. La première démarche dans l'étape du processus est l'administration de prétests, si ceux-ci sont appropriés. Les décisions, qui sont fondées sur les résultats des prétests, peuvent être prises à l'égard de la justesse des objectifs déjà déterminés. Après le test initial, les stratégies et les activités prévues sont réalisées dans la séquence préétablie. Les données collectées à cette étape permettent de vérifier si tout se déroule comme prévu et si les stratégies et les activités sont efficaces. Les buts fondamentaux de cette étape sont de voir si l'effort est déployé comme c'était envisagé, de déterminer le degré d'atteinte des objectifs du processus et de trouver des façons qui apportent des améliorations. L'étape du processus est désignée sous le nom d'évaluation formative.

Étape du produit

L'étape du produit nécessite des décisions prises à la fin ou probablement à la fin d'un cycle de pédagogie, d'un programme ou d'un projet. Les décisions prises durant l'étape du produit sont fondées sur les résultats des postests et sur d'autres types cumulatifs de données. Le but principal de cette étape de produit est de collecter des données afin de prendre des décisions sur l'efficacité générale de la pédagogie, d'un programme ou d'un projet. Durant cette étape, on détermine si les objectifs connexes au produit ont été atteints et si oui, dans quelle mesure ils ont été atteints. Après l'analyse et l'interprétation de données, il y a la préparation d'un rapport qui décrit les objectifs, la procédure et les résultats de l'effort. Les résultats de l'étape du produit de l'évaluation sont au moins employés de deux principales manières :

- i. ils fournissent de la rétroaction et servent de repères à tous ceux qui ont participé à l'effort
- ii. ils fournissent de la rétroaction aux décideurs à l'extérieur, comme les parents, les directeurs, les membres du conseil scolaire et les aides publiques

Les résultats de l'étape du produit doivent être interprétés avec prudence. Par exemple, le non-respect des objectifs n'est pas nécessairement désastreux; le degré d'atteinte doit être considéré. L'étape du produit de l'évaluation est appelée l'évaluation sommative.



Évaluation formative

Activité 1

Après avoir lu la note ci-dessus sur les différents types d'évaluation, remplissez le tableau ci-dessous.

Type d'évaluation	Principales variables mesurées	Principaux types de décisions	Quel est le type d'évaluation important pour l'éducation?
Élève			
Programme éducatif			
École			
Population dense			
Programmes spéciaux			
Personnel			

Activité 2

Soulevez cinq problèmes liés à l'évaluation de la fin.

- 1 _____
- 2 _____
- 3 _____
- 4 _____
- 5 _____

Activité 3

Remplissez le tableau ci-dessous.

Étape de l'évaluation	Exemples de décisions	Activités particulières
Planification		
Processus		
Produit		



Évaluation - Étude de cas

Choisissez une des écoles de votre région et interviewez le directeur ou la directrice de l'école ou quiconque chargé des examens de l'école à propos de la procédure à suivre lors des examens.

Les questions suivantes serviront de guide.

- 1 Que signifie une évaluation pour cette école? Qu'est-ce qui la motive?
- 2 Quels types d'évaluation sont utilisés dans cette école? Pourquoi est-elle utilisée?
- 3 Comment l'école planifie-t-elle et met-elle en oeuvre l'évaluation de l'élève? Quelle est la démarche à suivre pour chaque étape? Comment les résultats sont-ils utilisés? Qui a accès aux résultats de l'évaluation et comment sont-ils utilisés?

Évaluation des projets d'enseignement

L'évaluation des projets d'enseignement fait partie intégrante du processus de revue et d'évaluation des systèmes d'éducation.

La lecture obligatoire pour l'étude #1 (Designing Evaluation for Education Projects) et la rédaction d'un compte rendu sur l'évaluation des projets d'enseignement.

- 1 Qu'est-ce qu'une évaluation de projet?
- 2 Pourquoi une évaluation de projet est-elle importante en éducation?
- 3 Quelles sont les étapes à suivre dans une évaluation de projet?
- 4 Quels sont les différents types d'évaluation de projet?
- 5 Quels sont les différents instruments pour collecter les données utilisés dans l'évaluation de projet?

Votre compte rendu devrait être tapé et avoir entre 1000 et 2000 mots.

Assurez-vous d'avoir utilisé vos propres mots dans votre résumé.



Activité d'apprentissage #2

CARACTÉRISTIQUES DES OBJECTIFS

Résumé

Les objectifs se trouvent au cœur des activités d'évaluation en éducation et guident le processus de l'éducation. Ils nous aident à déterminer ce qu'il faut mesurer et comment le faire. Ils montrent si le processus d'éducation a été une réussite, et ce qui peut être fait pour qu'il soit une réussite. Dans cette activité, les apprenants seront capables de connaître les objectifs et leur importance en éducation. De plus, ils maîtriseront les types d'objectifs éducationnels et leurs éléments. La taxonomie des objectifs pédagogiques de Bloom sera abordée et servira de guide pour vérifier si les objectifs que nous fixons et mesurons sont assez exhaustifs pour comporter tous les aspects de l'éducation.

Liste de concepts-clé

Les objectifs sont des énoncés spécifiques de ce qui doit être accompli et de la mesure dans laquelle cela doit être accompli. Ils doivent être mesurables et quantifiables.

L'évaluation des besoins est le procédé par lequel les besoins sont identifiés et leur priorité, déterminée.

Les objectifs axés sur les compétences ont pour but de communiquer les comportements jugés nécessaires au succès futur des étudiants en incitant ces derniers à les adopter durant leurs études.

Les objectifs de maîtrise sont des objectifs que devraient atteindre tous les élèves, peu importe leurs aptitudes ou leurs acquis.

La taxinomie des objectifs d'apprentissage a pour but de classer les objectifs selon leur complexité présumée.

Les habiletés affectives sont liées à des objectifs intangibles dont la promotion et la mesure sont difficiles; par exemple les sentiments, les attitudes, les intérêts et les valeurs.

Les habiletés psychomotrices nécessitent des habiletés physiques, dont la coordination musculaire, les habiletés motrices, la manipulation d'objets ou la coordination neuromusculaire.

Buts : les buts sont des énoncés généraux portant sur des résultats à long terme.

Les objectifs intermédiaires sont plus précis que des objectifs généraux, mais pas aussi détaillés que des objectifs spécifiques.

Les objectifs spécifiques ont trait aux résultats les plus distincts.



Contenu

La nature des objectifs

Les objectifs sont des énoncés spécifiques de ce qui doit être accompli et de la mesure dans laquelle cela doit être accompli. Ils doivent être mesurables et quantifiables. Ils sont réalisés avant le déploiement d'un effort et peuvent faire l'objet d'un examen technique et approfondi. Un objectif pédagogique est un but exprimé par un énoncé décrivant un changement proposé chez l'apprenant- un énoncé qui expliquerait ce que l'apprenant devrait être après avoir terminé une expérience d'apprentissage avec succès. L'apprentissage est un processus inféré qui n'est pas directement observable; ainsi, on doit déterminer le comportement ou le rendement

qui constituera une preuve suffisante que l'aptitude recherchée existe. Que la capacité soit acquise ou apprise, à la suite de l'enseignement, elle ne peut qu'être déterminée en observant le rendement à deux différents moments : avant et après l'enseignement, et en vérifiant s'il y a eu un changement dans la réalisation. L'apprentissage en soi ne peut être observé ou mesuré directement, mais les changements dans le rendement peuvent l'être.

Les objectifs guident les activités d'un effort. Ils établissent une stratégie et des activités générales qui favorisent leur accomplissement. Conséquemment, des ensembles différents d'objectifs posséderont différentes stratégies et activités. Certains soutiennent que les objectifs donnent lieu à une procédure, des normes et des comportements rigides et contraignants qui ne laissent pas libre cours à la créativité et la spontanéité et qui ne réussissent pas à tenir compte des différences individuelles. Ils mettent également en doute l'objectivation de tout résultat pédagogique et un de leur argument est que nous avons tendance à fixer des objectifs faciles à mesurer qui ne sont pas nécessairement importants. Cependant, il faut noter que la plupart des résultats voulus peuvent être observables et quantifiables.

Évaluation des besoins

Pour le système d'éducation, ce serait quasiment impossible de tenter d'encourager la réalisation de tous les objectifs possibles; formellement ou informellement, chaque effort pédagogique devrait choisir et placer en priorité les buts et les objectifs qui seront concernés. Conceptuellement, s'il y a un écart entre la façon dont les choses sont (statut actuel ou statu quo) et celle dont nous aimerions qu'elles soient (le résultat voulu), il y a un *besoin*. Tout dépendant du caractère spécifique



de la définition du besoin, il peut être exprimé en un but ou objectif; l'intention étant d'éliminer le besoin. L'évaluation des besoins est le procédé par lequel les besoins sont identifiés et leur priorité, déterminée. Les buts et les objectifs fondamentaux du système d'éducation sont demeurés relativement stables, mais leur importance et leurs priorités ont changé au fil du temps. L'évaluation des besoins peut s'appliquer au système éducatif, notamment aux classes, aux programmes, aux projets, aux écoles, aux systèmes scolaires, aux États et aux pays, et peut être utilisée à tous les niveaux et dans tous les aspects.

Classification et formulation des objectifs

Types d'objectifs

Les objectifs sont pareils, peu importe le type d'évaluation, et tous les objectifs indiquent des résultats voulus. Une différence fondamentale est que certains types d'évaluation, et les objectifs qui leur correspondent, traitent surtout des réalisations individuelles, tandis que les autres traitent surtout du rendement des groupes.

Littéralement, des milliers et des milliers d'objectifs couvrant pratiquement chaque domaine d'enseignement ont été écrits pour énoncer les comportements voulus des élèves. Par contre, très peu d'objectifs existent sur le plan des objectifs de rendement pour les professeurs, et il y en a encore moins pour les gestionnaires, et pratiquement aucun pour le programme spécial et le personnel du projet; cela s'explique principalement par le manque de preuve empirique reliant les comportements donnés et la réussite dans l'emploi.

a) Objectifs axés vers la compétence

Les élèves de divers programmes doivent réaliser des objectifs ou des compétences qui sont généralement conçus par des professeurs d'université ou d'établissement d'enseignement supérieur, des experts dans leurs champs respectifs, et qui sont passés en revue par l'école privée ou les groupes communautaires appropriés. Les objectifs axés vers la compétence visent la communication de comportements que l'on estime être nécessaires pour un rendement de travail réussi et la démonstration de ces comportements pendant les études.

b) Maîtrise des objectifs

On estime que certains objectifs d'enseignement sont nécessaires, les notions fondamentales de base, c.-à-d. les objectifs qui, pratiquement, devraient être atteints par tous les élèves, quels que soient leurs acquis et leurs compétences; ils



sont appelés la maîtrise des objectifs ou les compétences et les objectifs minimaux. Dans ce système de maîtrise, la quantité de notions apprises, comme c'est indiqué dans l'objectif, est constante; l'enseignement est personnalisé et le temps requis pour atteindre ces objectifs varie d'un élève à l'autre.

Taxonomie des objectifs pédagogiques

Les taxonomies classent tous les objectifs dans une hiérarchie de catégories selon le niveau de complexité supposé. Chaque catégorie à venir implique que les comportements soient prétendument plus complexes que le comportement précédent et on estime que chacun est un préalable au suivant. Le but énoncé des taxonomies est de favoriser la communication. L'existence des taxonomies attire aussi l'attention sur de nombreux résultats comportementaux; de la sorte, il y aura moins de chances que tous les objectifs pour un effort donné fassent intervenir uniquement des comportements du niveau inférieur. On croit que tout objectif d'enseignement peut être classé dans l'une des trois taxonomies ou dans l'un des trois domaines – cognitif, affectif et psychomoteur.

1. Le domaine cognitif

Sans aucun doute, la taxonomie des objectifs cognitifs a sensibilisé les éducateurs à un large éventail de compétences intervenant dans l'apprentissage cognitif. Chacune des six principales catégories représente une sorte différente du processus d'apprentissage. Les principales catégories de la taxonomie cognitive des objectifs d'enseignement sont :

- a) la connaissance;
- b) la compréhension ;
- c) l'application ;
- d) l'analyse;
- e) la synthèse ;
- f) l'évaluation.

Robert Gagné a abordé la classification des types d'apprentissage d'une perspective différente. Il croit que les types ou les classes d'apprentissage peuvent être classés selon les critères nécessaires pour faciliter l'apprentissage de chaque type. Gagne a distingué huit classes d'apprentissage et des conditions d'enseignement correspondantes pour l'apprentissage qui leur sont rattachées a)l'apprentissage de signe b)l'apprentissage par stimulus et réponse c)l'enchaînement d)l'association verbale e)l'apprentissage par discrimination f)l'apprentissage de concept et g) la résolution de problèmes.



Antérieurement, les éducateurs ont été critiqués d'avoir porté une trop grande attention aux résultats cognitifs et d'avoir négligé les sentiments et les attitudes des élèves. Le domaine affectif traite de tels résultats.

2. Le domaine affectif

La taxonomie des objectifs affectifs s'occupe de résultats qui sont beaucoup plus difficiles à encourager et à mesurer, des impondérables comme les sentiments, les attitudes, les intérêts et les valeurs. Les catégories affectives représentent une hiérarchie d'assimilation qui va de la volonté à recevoir, ou l'assistance, à la caractérisation par une valeur. Comme avec tous les objectifs, le but est d'identifier des comportements observables et quantifiables à partir desquels nous pouvons inférer l'apprentissage; manifestation, c'est plus difficile pour les résultats affectifs. La réalisation des objectifs affectifs peut être déterminée par l'administration d'instruments d'auto-évaluation qui entraînent une quantité considérable de problèmes. Les principales catégories de la taxonomie affective des objectifs pédagogiques sont :

- a) la réception;
- b) la réponse;
- c) l'évaluation;
- d) l'organisation;
- e) la caractérisation par valeur.

3. Le domaine psychomoteur

Le domaine psychomoteur comporte les capacités physiques, plus particulièrement celles qui font intervenir les capacités musculaires ou les habiletés motrices, la manipulation d'objectifs, ou la coordination neuromusculaire. La tâche pour le développement de la taxonomie dans le domaine psychomoteur est plus complexe que pour les autres domaines. Bien qu'il n'y ait pas de catégories largement acceptées pour ce domaine, les catégories suivantes sont utilisées dans une partie de la documentation sur le sujet :

- a) les réflexes
- b) les mouvements fondamentaux
- c) les habiletés perceptives
- d) les capacités physiques
- e) les mouvements adroits
- f) la communication gestuelle



Niveaux de spécificité

Les objectifs peuvent être écrits à divers niveaux de spécificité. À un bout de la gamme, il y a des énoncés très généraux sur les résultats à long terme; d'ordinaire, on les appelle les buts ou les objectifs généraux. À l'autre bout de la gamme, il y a d'autres énoncés très précis sur des résultats plus immédiats, chacun représentant un comportement ou un résultat; ils sont généralement appelés les objectifs spécifiques.

Buts/ objectifs généraux

Les buts sont des énoncés généraux sur les résultats à long terme. L'atteinte de ces buts ou ces objectifs généraux ne peut pas être mesurée directement. Les buts de l'éducation ont tendance à être établis à l'échelon national.

Objectifs intermédiaires

Les buts sont fréquemment traduits en objectifs intermédiaires. À ce niveau, les objectifs sont plus précis que des objectifs généraux, mais pas aussi détaillés que des objectifs spécifiques. Dans plusieurs situations, on les appelle les objectifs cumulatifs, les objectifs d'un cours et les tâches. D'ordinaire, de tels objectifs sont mesurables et représentent normalement des comportements complexes prévus à la suite de la réalisation d'un certain nombre d'objectifs spécifiques préalables.

Objectifs spécifiques

Les objectifs spécifiques ont trait aux résultats les plus distincts. Ils sont des résultats quantifiables et constituent les bases d'activités quotidiennes comme l'enseignement. La procédure d'analyse de tâches est une méthode qui permet de convertir les objectifs cumulatifs en objectifs spécifiques.

Le tableau ci-dessous présente un résumé de la terminologie liée au niveau de spécificité des objectifs.



Niveaux	Définitions	Termes connexes
Objectifs généraux	Énoncés généraux sur les résultats à long terme	Buts
Objectifs intermédiaires	Énoncés plus précis sur plus de résultats à court terme	Objectifs du deuxième niveau Objectifs du module Objectifs du cours Objectifs terminaux Objectifs d'unité
Objectifs spécifiques	Énoncés précis sur des résultats immédiats	Objectifs pédagogiques Objectifs comportementaux

Un objectif peut être vague à un tel point qu'il est ambigu ou détaillé à un tel point qu'il est insignifiant et peu réaliste. Un principe directeur raisonnable implique que chaque objectif devrait représenter une tâche distincte. Une tâche peut être définie comme le plus petit aspect du rendement ayant un but distinct et indépendant. Plus un objectif est formulé simplement, plus il est clairement pratique.

Objectifs exposés

Le choix ou l'élaboration d'un objectif est sans contredit la partie la plus importante de tout programme éducatif. L'objectif guide tous les autres aspects du programme, y compris le contenu et la méthodologie qui seront utilisés pour l'enseignement et l'évaluation. Avant tout, nous devons évaluer la spécificité de l'objectif. L'objectif devrait communiquer le résultat escompté assez clairement pour que tout éducateur l'évalue sensiblement de la même façon et qu'il soit en mesure de dire si l'objectif a été atteint.



Éléments des objectifs

a) Comportement

Dans un objectif, le comportement indique ce qui se produira, ou, comme c'est le cas pour les objectifs pédagogiques, quel rendement observable et mesurable constituera une preuve satisfaisante que l'apprentissage voulu s'est produit. Un objectif devrait exposer le résultat d'apprentissage prévu par un verbe d'action représentant un résultat observable. Chaque objectif devrait comporter la réalisation d'un seul comportement, sinon, l'évaluation du rendement est confondue avec d'autres objectifs.

b) Condition

Les conditions pour un objectif sont « données », c.-à-d. le contexte dans lequel le comportement sera manifesté. D'ordinaire, elles précisent le matériel nécessaire, les ressources permises, ou les restrictions imposées, et le mot « données » n'est pas nécessairement employé.

c) Critère

Lorsqu'il est formulé, le critère d'un objectif est l'aspect qui nous indique la bonne façon dont un comportement doit être réalisé. Il est la norme qui permet de comparer le rendement actuel. Si le critère est évident ou compris à 100 %, il n'est pas exposé généralement.



Évaluation formative

1. Les lectures obligatoires pour les études #2 et #3. Pour chacun des comportements suivants, cochez « C » si c'est un comportement cognitif et un comportement affectif. Aussi, cochez « S » s'il représente des résultats d'ordre supérieur et « I » s'il représente des résultats d'ordre inférieur.

No	Comportement	Type		Niveau	
		C	A	S	I
1.	Récite l'article I de la Constitution				
2.	Équilibre les équations chimiques				
3.	Porte attention au film sur la génétique				
4.	Interprète un graphique				
5.	N'achète que des aliments halal				
6.	Définit le terme « évaluation » en se servant de ses propres mots				
7.	Écoute une station de radio classique quotidiennement en route vers le travail et en revenant du travail				
8.	Participe à un concours d'orthographe				
9.	Prépare une déclaration fiscale				
10.	Connait la signification des panneaux de signalisation				



2. Pour chacun des énoncés suivants, cochez « B » si c'est un but, « I » si c'est un objectif intermédiaire ou « S » si c'est un objectif spécifique.

No	Énoncé	B	I	S
1.	Devenir un bon citoyen			
2.	Taper une lettre commerciale dans la forme correcte			
3.	Indiquer trois sources alimentaires de fer			
4.	Réussir trois paniers sur dix essais en lançant de la ligne des lancers francs			
5.	Maîtriser les compétences fondamentales			
6.	Écrire des phrases grammaticalement correctes			
7.	Devenir autosuffisant sur le plan économique			
8.	Effectuer des opérations de base sur les fractions			
9.	Être capable de dresser une liste de romans du 20e siècle et d'indiquer l'auteur de chacun avec une précision de 80 %			
10.	Emprunter de son propre gré deux livres de la bibliothèque durant la session			



3. Étudiez le contenu des lectures #4 et #5 et effectuez l'exercice suivant.

1. Choisissez deux des sujets suivants enseignés dans un programme de cours d'école secondaire :
 - a. la périodicité des produits chimiques;
 - b. l'appareil génital ;
 - c. la force de gravitation;
 - d. les roches;
 - e. les tremblements de terre;
 - f. les guerres mondiales.

2. Pour chacun des sujets choisis, écrivez au moins deux objectifs généraux et environ quatre objectifs spécifiques que les apprenants doivent atteindre à la fin de l'évaluation.

3. Indiquez le niveau d'enseignement secondaire pour lequel ces objectifs sont adéquats.

4. Pour chacun des objectifs élaborés, classez-les en fonction de leur domaine connexe – cognitif, affectif ou psychomoteur.

5. Pour chacun des objectifs élaborés, indiquez de nouveau leurs aspects comme les comportements, les conditions et les critères.



Activité d'apprentissage #3

CLASSEMENT ET CHOIX DES TESTS

Résumé

Cette activité d'apprentissage vise la maîtrise de certains concepts par les élèves dans le domaine de l'évaluation et des tests en éducation. L'activité commence par certains concepts de base liés à la mesure et aux tests. Ensuite, les modes de collecte de données employés pour l'évaluation sont examinés brièvement. En détail, les systèmes de classification utilisés pour classer différents tests. Les caractéristiques les plus importantes d'un bon test sont exposées brièvement plus loin dans cette section du module. Finalement, vous aurez accès à certains éléments concernant le processus du choix d'un test (généralement standardisé) qui proviennent d'un certain nombre d'options.

Liste de concepts-clé

Un compte rendu comprend un certain nombre limité de questions et les répondants doivent rédiger les réponses.

Une épreuve objective est un test dans lequel la subjectivité dans la correction est éliminée, du moins théoriquement.

Un test standardisé est un test élaboré par des spécialistes de la matière et de l'évaluation.

Un test conçu localement est généralement conçu localement par les professeurs pour un but particulier.

Des données sur l'auto-évaluation sont des réponses orales ou écrites.

L'échelle de notation est un instrument qui comporte plusieurs items liés à une variable donnée; chaque élément représente un continuum de catégories entre deux extrêmes.

Les tests de connaissances sont des tests qui mesurent le statut d'un individu en fonction de ses aptitudes dans des domaines de connaissance divers.

Les tests de caractère ou de personnalité sont conçus pour évaluer les caractéristiques des personnes ainsi qu'un certain nombre de qualités.

Les tests d'aptitude sont des façons de mesurer le potentiel employées pour prédire le rendement probable d'une personne dans une situation ultérieure.



Les normes de rendement sont les critères servant à comparer les résultats des évaluations afin de les interpréter.

Une évaluation normative est tout test, standardisé ou conçu localement, qui présente et interprète chaque note en fonction de sa position par rapport aux autres notes obtenues dans le même test.

Un test critique est tout test qui présente et interprète chaque note par rapport à un standard absolu.

L'objectivité du test est le fait que, peu importe le correcteur, la note du test reste sensiblement la même.

La discrimination est le fait qu'un test détecte ou mesure les petites variations de rendement ou de réalisation.

La représentativité indique à quel degré un test couvre les principaux objectifs du cours qui doivent être mesurés.

La validité indique ce qu'un test est censé mesurer.

La validité du contenu est la mesure dans laquelle les items d'un test correspondent à l'étendue des items qu'un test devrait normalement couvrir.

La validité critique est la mesure dans laquelle un test réussit effectivement à prédire des critères ou à indiquer une construction.

La validité concurrente est rencontrée lorsque la mesure des critères a lieu en même temps que l'obtention des notes.

La validité prédictive est rencontrée lorsque la mesure des critères a lieu après le test.

La validité conceptuelle est la mesure dans laquelle un test démontre un lien entre les résultats obtenus au test et le trait théorique prédit.

La fiabilité est la cohérence des mesures. Un test est considéré fiable lorsque les mêmes résultats sont obtenus de manière répétée.



Contenu

Évaluation et test

Comme nous l'avons vu précédemment, l'évaluation est un aspect essentiel du processus d'évaluation. Elle joue un rôle essentiel puisque les décisions qui en découlent sont aussi bonnes que les données qui ont servi à cette prise de décision. De façon générale, la collecte de donnée est présente dans toutes les étapes de l'évaluation- l'étape de la planification, du processus et du produit. Toutefois, l'évaluation, qui est la mesure dans laquelle quelqu'un ou quelque chose possède un trait donné, se produit généralement dans l'étape du processus et celle du produit.

Le test est nécessaire à certains endroits et utile à d'autres. Il peut être effectué à la fin d'un cycle d'instruction –semestre, session ou unité. De tels postests servent à déterminer dans quelle mesure les objectifs (formels ou informels) ont été réalisés, qu'ils soient des objectifs pédagogiques ou des objectifs de programme. Les données sur les prétests ou les données de références sont fréquemment collectées au début du cycle. Les prétests ont plusieurs utilités, la plus importante étant que la connaissance de l'état actuel d'un groupe peut guider les activités ultérieures et constituer une base de comparaison pour les résultats des postests.

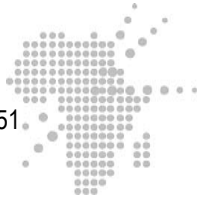
Le test est utile dans une panoplie de situations. Un professeur peut administrer des tests

pour les comportements d'entrée pour déterminer si les préalables hypothétiques ont été effectivement obtenus. Un projet spécial conçu pour réduire le décrochage peut employer des tests d'attitude et des tests de variables de personnalité comme l'introversiion, l'agressivité et l'anxiété pour essayer d'identifier les décrocheurs potentiels et de mieux comprendre les élèves ayant des difficultés. Une école peut administrer des tests d'aptitudes aux études afin de déterminer des objectifs d'accomplissement pour les élèves et de participer au processus d'orientation.

Collecte de données

Il y a trois principales façons de collecter des données

- i) employer un instrument standardisé
- ii) employer un instrument conçu localement
- iii) consigner les données accessibles naturellement (comme les moyennes pondérées cumulatives et l'absentéisme)



Tout dépendant de la situation, une de ces façons peut être très adéquate ou une combinaison peut être exigée. La collecte des données accessibles, qui nécessite un minimum d'efforts, semble très attrayante. Cependant, ces types de données conviennent à peu de situations. Même lorsqu'ils sont appropriés –c'est-à-dire lorsqu'ils facilitent la prise de décision prévue- il y a des problèmes inhérents à ce type de données. Par exemple, la même cote graphique ne représente pas nécessairement le même niveau de rendement, même pour deux différentes classes dans la même école ou deux écoles différentes dans le même système. De plus, les dossiers, pour lesquels les données sont collectées, peuvent être incomplets et confus. Élaborer un instrument pour un but spécifique comporte aussi plusieurs inconvénients importants. L'élaboration d'un « bon » instrument nécessite beaucoup de temps et d'efforts, et des compétences particulières. Pour acquérir les compétences nécessaires pour la mise au point d'un bon instrument, il faut une formation qui équivaut au moins à un cours sur les tests et l'évaluation.

En revanche, le temps requis pour choisir un instrument adéquat (d'ordinaire, ce choix est fait parmi les instruments standardisés et offerts sur le marché) est moindre que celui nécessaire pour concevoir un instrument qui mesure la même chose. D'autres instruments standardisés sont généralement conçus par des experts possédant les compétences requises. Des milliers d'instruments standardisés sont offerts et donnent accès à un large éventail de données à diverses fins. De nombreux instruments de mesure ont été conçus pour les principaux domaines qui portent sur le rendement, la personnalité et l'aptitude. Chacune de ces catégories peut être à son tour divisée en sous-catégories. En général, faire des recherches pour savoir si un instrument adéquat est déjà offert avant d'investir dans la conception d'un instrument est une bonne idée. Toutefois, il existe des situations où l'emploi des instruments offerts s'avère peu réaliste ou inadéquat. Par exemple, un test élaboré par le professeur est plus approprié ou valable pour déterminer dans quelle mesure les élèves ont réalisé les objectifs pour une unité donnée.



Systèmes de classification

À ce stade-ci, on devrait insister sur le fait qu'un test n'est pas nécessairement un ensemble écrit de questions auxquelles une personne répond afin de déterminer si elle réussit. Une définition plus inclusive d'un test est le « moyen d'évaluer les connaissances, les compétences, l'intelligence ou l'aptitude d'une personne ou d'un groupe ». Les tests donnent des résultats numériques qui peuvent être utilisés pour identifier, classer ou évaluer ceux qui le font. Bien qu'en pratique, la plupart des tests emploient la technique « papier et crayon », il existe différentes sortes de tests et de nombreuses façons pour les classer. Les systèmes de classification variés se recouvrent beaucoup et les catégories sont loin de s'exclure mutuellement. Tout test peut être classé dans plus d'un secteur.

Comportements par réponse

Le terme « comportements par réponse » se rapporte à la façon dont les comportements qui doivent être mesurés sont montrés. Bien que dans certains cas, les réponses aux questions ou à d'autres stimuli soient données oralement, elles sont généralement écrites ou prennent la forme d'un rendement réel.

A. Réponses écrites

Les tests écrits peuvent être classés en compte-rendu (subjectif) ou en test standardisé ou conçu localement (objectif).

Compte-rendu par rapport aux tests objectifs

i) Compte-rendu

Un compte-rendu comporte un nombre de questions limité et les répondants doivent rédiger leurs réponses, généralement assez longues. Un exemple de question est « Identifiez les principales réformes de l'Afrique en éducation lors de la période coloniale et commentez. » Déterminer la justesse des réponses à de telles questions comporte une certaine part de subjectivité

ii) Tests objectifs

Un test objectif est un test où il n'y a aucune part de subjectivité dans la notation, du moins théoriquement. En d'autres termes, si deux personnes font la notation d'un test donné, elles devraient arriver au même résultat. Il existe plusieurs sortes de tests objectifs : tests à choix multiples, vrai ou faux, tests d'appariement et tests à court développement.



Tests standardisés par rapport aux tests conçus localement

i) Un test standardisé est élaboré par des spécialistes de la matière et de l'évaluation et est vérifié dans la pratique conformément à une procédure administrative uniforme. Ce test est aussi révisé pour satisfaire à certains critères et corrigé et interprété selon une procédure et des normes uniformes. La standardisation couvre tous les aspects du test pour faire en sorte qu'il peut être administré et noté exactement de la même façon chaque fois qu'il est donné. Bien que les autres instruments de mesure puissent être standardisés, la plupart des tests standardisés sont objectifs et les tests écrits nécessitent des réponses écrites. Malgré certaines exceptions, la très grande majorité des tests standardisés a été administrée à des groupes qui sont appelés le groupe de référence. Les résultats du groupe de référence pour un test donné constituent une base de comparaison et d'interprétation pour les autres groupes qui subiront le test. Les résultats d'un groupe de référence sont appelés les normes. L'idéal serait que le groupe de référence soit un grand groupe bien défini et représentatif du groupe et des sous-groupes pour lesquels le test est conçu.

ii) Test conçu localement

L'opposé du test standardisé est évidemment le test non standardisé. D'ordinaire, de tels tests sont conçus localement pour un but précis. Les tests employés en classe par des professeurs sont un exemple de tests conçus localement. Ces tests ne possèdent pas les caractéristiques des tests standardisés. Un test élaboré localement peut être aussi bon qu'un test standardisé, mais cela n'arrive pas souvent. L'utilisation de test conçu localement est généralement plus pratique et plus adéquate. Si l'on compare les tests standardisés aux tests conçus localement, les tests conçus localement refléteraient mieux ce qui a été enseigné en classe.

B. Tests de rendement

Pour de nombreux objectifs et domaines d'apprentissage, l'utilisation de tests écrits est une façon inadéquate de mesurer le comportement. Par exemple, on ne peut déterminer dans quelle mesure un élève peut bien écrire une lettre avec un test à choix multiples ou une question ouverte. Le rendement est l'un de ces domaines. Il peut prendre la forme d'une procédure ou d'un produit. Une procédure est une série d'étapes, généralement dans un ordre précis, exécutées dans l'accomplissement d'un acte ou d'une tâche. Des exemples de procédure comprennent : le réglage d'un microscope, la passe d'un ballon de football, le réglage de marges sur une machine à écrire, dessiner des figures géométriques ou



calculer une somme de chiffres avec Excel. Un produit est un résultat tangible. Une lettre tapée à la machine, une peinture, un poème et un projet en sciences sont des exemples de produit. Dans chaque cas, le rendement est observé et évalué dans un certain sens. Ainsi, le rendement nécessite l'exécution d'un acte ou le développement d'un produit afin de déterminer si une aptitude ou un trait donné existe, et dans quelle mesure.

Méthodes pour collecter les données

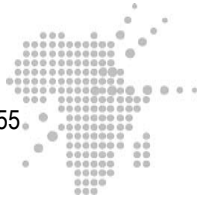
Il existe de nombreuses façons de collecter les données et classer les méthodes de collecte de données n'est pas une mince tâche. Toutefois, une manière logique de les classer au départ est de vérifier si les données sont obtenues par l'intermédiaire d'une auto-évaluation ou d'une observation.

a) Auto-évaluation

Les données d'auto-évaluation sont composées de réponses orales ou écrites provenant des personnes. Les données issues de l'administration de tests écrits standardisés ou conçus localement (cela comprend certains tests de rendement, de personnalité et d'aptitude) constituent un bon type de données d'auto-évaluation. Un autre type d'instrument d'auto-évaluation employé dans certains efforts d'évaluation est le questionnaire, que vous connaissez bien probablement. Aussi, les entretiens sont parfois utilisés.

b) Observation

Lorsqu'on utilise l'observation, les données sont collectées par observation et non pas à la suite de questions et de réponses. Généralement, une personne observée ne prend rien en note; elle fait quelque chose et ce comportement est observé et noté. Pour certaines questions d'évaluation, l'observation est manifestement l'approche la plus adéquate. Par exemple, vous pourriez demander aux élèves de parler de leur sportivité et demander aux professeurs comment ils gèrent les problèmes de comportement, mais en fait, des renseignements plus objectifs seraient probablement obtenus par l'observation des élèves à des manifestations sportives et l'observation des professeurs dans leur classe. L'observation naturelle et l'observation de simulation sont deux genres d'observation utilisés dans les efforts d'évaluation. Certains types de comportement peuvent uniquement être observables, car ils surviennent naturellement. Pour de telles situations, l'observateur ne maîtrise et ne manipule rien et déploie beaucoup d'efforts pour ne pas influencer le déroulement de l'observation de quelque façon. Par exemple, l'observation est le meilleur moyen d'aborder le comportement en classe. Pour l'observation de simulation, l'évaluateur reproduit la situation qui doit être observée et dit aux participants à quelles activités ils se livrent. Cette technique



permet à l'évaluateur d'observer des comportements qui se produisent rarement ou jamais dans des situations naturelles.

c) L'échelle de notation

Elle est un instrument qui comporte plusieurs éléments liés à une variable donnée; chaque élément représente un continuum de catégories entre deux extrêmes; les personnes qui répondent aux questions mettent une marque pour indiquer leur position sur chaque question. Les échelles de notation peuvent être utilisées comme instruments pour l'auto-évaluation ou l'observation, selon le but pour lesquelles elles sont utilisées.

Comportements mesurés

Presque tous les comportements possibles qui peuvent être mesurés se retrouvent dans l'une des trois principales catégories : rendement, caractère et personnalité, et aptitude. Chacune d'elle peut être standardisée ou conçue localement. Ces catégories s'appliquent tout aussi bien aux trois domaines de résultat pédagogique : cognitif, affectif et psychomoteur.

1. Rendement

Le test de rendement évalue l'état correct des personnes par rapport au niveau dans des branches données de la connaissance ou des compétences. Il est adéquat pour de nombreux types d'évaluation en plus de l'évaluation individuelle de l'élève. Le test de rendement peut être standardisé (il est alors conçu pour couvrir de la matière commune à de nombreux cours du même type) ou conçu localement (préparé pour évaluer un ensemble précis de résultats d'apprentissage, établis par un professeur en particulier). Les tests standardisés sont accessibles pour les domaines individuels du programme éducatif ou prennent la forme de batteries, qui évaluent le rendement dans plusieurs domaines différents.

Un test de diagnostic est un type de test de rendement qui donne des résultats multiples pour chaque domaine de rendement; ces résultats facilitent l'identification de domaines précis, de défaut ou de difficulté d'apprentissage. Les éléments d'un test de diagnostic visent à identifier les compétences et les connaissances que les élèves doivent avoir acquises avant le prochain niveau. L'idéal serait que le diagnostic soit un processus continu et le professeur doit concevoir les tests de façon à ce qu'ils l'aident à découvrir les problèmes rencontrés par les élèves lors de leur cheminement dans le processus d'apprentissage.



2. Caractère et personnalité

Les tests de caractère et de personnalité sont conçus pour évaluer les caractéristiques des personnes ainsi qu'un certain nombre de qualités et pour évaluer les sentiments et les attitudes à l'égard de soi, des autres et diverses activités, institutions et situations. La plupart des tests de caractère et de personnalité sont des instruments d'auto-évaluation et vous pouvez demander aux personnes de répondre à une série de questions ou d'énoncés. Dans cette catégorie, il y a des instruments conçus pour évaluer la personnalité, les attitudes, la créativité, et l'intérêt des élèves.

3. Aptitude

Les tests d'aptitude sont des façons de mesurer le potentiel. Elles sont employées pour prédire le rendement probable d'une personne dans une situation ultérieure. Les tests d'aptitude générale sont aussi appelés de différentes manières : tests d'aptitude aux études, tests d'intelligence et tests d'efficience intellectuelle. Les tests d'aptitude permettent aussi de prédire le niveau de rendement probable d'une personne après qu'elle ait reçu certaines instructions ou formations précises. Ils se présentent sous forme de batteries ou de test individuel sur une matière ou un sujet précis. Bien que pratiquement tous les tests d'aptitude soient standardisés et donnés dans le cadre d'un programme de test scolaire, les résultats sont utiles pour les professeurs, les conseillers et les gestionnaires. Les tests d'aptitude à l'état de préparation (ou tests pronostiques) sont donnés avant l'enseignement ou la formation dans un domaine précis afin de déterminer si un élève est prêt à recevoir un enseignement ou en profiter, et dans quelle mesure. Les tests d'aptitude à l'état de préparation, qui font partie des tests d'aptitude, comprennent généralement l'évaluation de variables comme la discrimination auditive, la discrimination visuelle et les habiletés motrices.

Normes de rendement

Les normes de rendement sont les critères qui permettent de comparer les résultats de l'évaluation afin de les interpréter. Un résultat tout seul ne veut rien dire. Si je vous dis que Ahmed a eu 18 bonnes réponses, qu'est-ce que cela vous dit à propos de son rendement? Absolument rien. Maintenant, si je vous dis que la moyenne est de 15, au moins vous savez qu'il a eu plus haut que la moyenne. À la place, si je vous dis qu'une note de 17 est requise pour un objectif de maîtrise, vous ne savez rien sur le rendement du reste de la classe, mais vous savez que Ahmed a atteint l'objectif de maîtrise. Nous pouvons interpréter le résultat à un test de deux façons, l'une étant par comparaison aux autres résultats des élèves de la classe (ce qui est une mesure normative) et l'autre étant par comparaison du résultat à un critère prédéterminé (ce qui est la mesure critérielle).

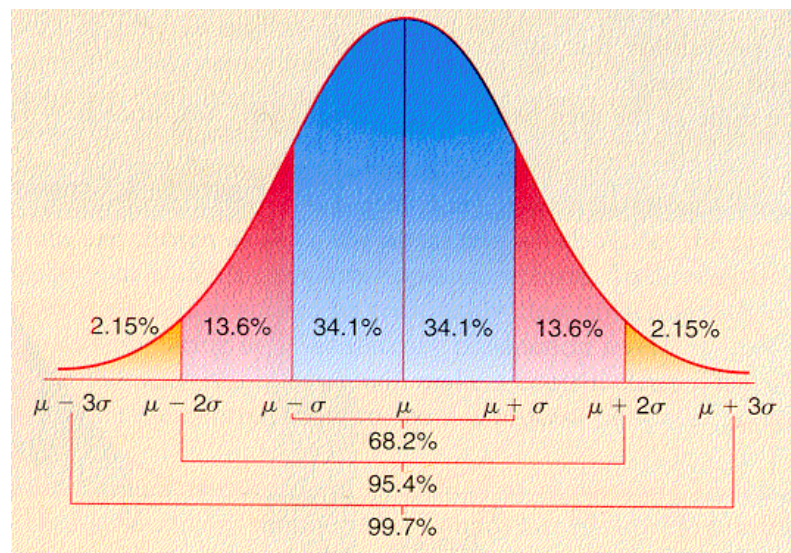


Standards normatifs

Tout test, standardisé ou conçu localement, qui présente et interprète chaque résultat en fonction de sa position relative par rapport aux autres notes obtenues dans le même test, est normatif. Par exemple, si quelqu'un a un QI total de 100, l'interprétation que l'on peut en faire est que l'intelligence mesurée est moyenne si l'on compare aux résultats obtenus par un groupe de référence. Les scores bruts résultant de l'administration de tests standardisés sont convertis en un autre indicateur qui montre la position relative. Le centile, que vous connaissez bien, s'apparente à une telle technique. Un centile donné indique le pourcentage des notes qui étaient plus basses que la note correspondante du centile. Par exemple, M. Ahmed aurait pu obtenir une note qui équivaldrait à un centile de 42 dans un test de mathématiques standardisé, ce qui signifie que 42 % des élèves ayant fait le test ont obtenu un résultat plus bas que Ahmed. Lorsque l'on communique les résultats de l'élève de cette façon, rien ne permet d'indiquer ce que Ahmed sait ou ne sait pas. Les seules interprétations que l'on peut en tirer concernent le rendement de Ahmed par rapport à celui des autres.

Les tests normatifs sont fondés sur l'idée que les traits évalués concernent les propriétés de la courbe normale. L'idée de la courbe normale est que les traits évalués existent en quantités différentes chez des gens différents. Certaines personnes en ont beaucoup, tandis que d'autres en ont peu, et la plupart des personnes en possèdent une certaine quantité, ce qui est appelé la quantité « moyenne ». Par exemple, si vous donnez un test de mathématiques à 100 élèves, étant donné qu'il est d'un niveau adéquat- c'est-à-dire pas trop facile ni trop difficile- une petite partie de la classe aura de bons résultats et une autre partie égale aura des résultats faibles, tandis que la majorité aura obtenu une note moyenne.

La courbe normale est illustrée ci-dessous.





La plupart des notes sont (la fréquence la plus élevée) moyennes, et peu de notes sont au-dessus ou en dessous de la moyenne, et très peu largement au-dessus ou en dessous de la moyenne. Le groupe ayant eu les notes moyennes comprend environ 65 % des résultats. La mesure normative utilisée dans cette distribution hypothétique de notes est appliquée à de vraies situations en classe et d'évaluation. C'est-à-dire, lorsque les meilleurs 2-3 % ont des A (peu importe leurs scores), les prochains 15 % obtiennent des B, les prochains 65 % reçoivent des C, les prochains 15 % ont des D, et les derniers 2-3 % reçoivent des F. Les notes en lettre sont attribuées sans tenir compte du rendement réel de l'élève.

Il y a un problème avec les standards normatifs. Certaines personnes peuvent obtenir de bonnes notes (ou inversement, de mauvaises notes) sans nécessairement afficher le niveau de rendement correspondant. Une telle méthode peut stimuler les élèves, mais elle peut aussi causer un problème particulier lorsque la classe est plus homogène.

Interprétations critériées

Tout test qui présente et interprète chaque résultat **en fonction** d'un standard absolu est critériel. En d'autres mots, l'interprétation du résultat d'une personne n'a rien à voir avec le résultat de quiconque. Un score est comparé avec une norme ou un rendement, et non avec les scores des autres. Lorsque le test critériel est utilisé, chaque personne qui le passe peut obtenir un résultat fort ou faible. Dans ce contexte, un critère peut être défini comme étant un domaine de comportements qui mesurent un objectif.

Caractéristiques d'un bon test

Certaines caractéristiques sont souhaitables pour tout test. Les tests standardisés, qui sont élaborés par des experts dans leur matière et en évaluation, sont le plus susceptibles de répondre à ces normes. Ces caractéristiques, qui sont abordées plus en détail, comprennent :

- a) l'objectivité du test
- b) la discrimination
- c) la perspective globale
- d) la validité
- e) la fiabilité
- f) la spécification des conditions dans lesquelles le test sera subi
- g) l'orientation de la notation et l'interprétation



Objectivité du test

L'objectivité du test est le fait que, peu importe le correcteur, la note du test reste sensiblement la même. Un test est objectif lorsque l'opinion, la partialité ou le jugement individuel ne constituent pas de facteurs déterminants dans la cotation. Les tests peuvent être effectués par plus d'une personne, à un moment différent ou simultanément. En éducation, nous nous demandons dans quelle mesure la cotation de ces correcteurs individuels est la même. Si la différence possible entre les corrections d'un même test est grande, ce test est peu objectif. Bien que les personnes puissent percevoir l'information de façon différente, naturellement, nous présumons que plus un test est objectif, plus il aspire à l'évaluation de haute qualité que nous souhaitons tous en éducation. Cela ne veut pas dire que les tests ne possédant pas un niveau élevé d'objectivité (comme les tests subjectifs) sont de mauvaise qualité. Même les tests subjectifs, bien qu'ils soient conçus pour évaluer des informations pouvant être abordées d'angles différents, une certaine part d'objectivité est nécessaire. Le concepteur individuel doit avoir quelque chose en tête qui constitue un bon rendement pour un tel test. Cette chose peut être multiple, mais les critères doivent être établis pour garantir que toute cotation soit assez juste pour contribuer à établir une discrimination entre les élèves sur une telle base. L'objectivité, c'est relatif.

Discrimination

Le test devrait être construit pour détecter ou évaluer les petites différences dans le rendement ou l'atteinte; c'est essentiel si le test doit être utilisé pour classer les élèves en fonction de leur rendement individuel ou pour attribuer des notes. Par contre, ce n'est pas un facteur important si le test est employé pour mesurer le niveau de la classe entière ou comme quiz pédagogique ayant comme but principal l'enseignement plutôt que l'évaluation. Comme il est vrai pour la validité, la fiabilité et l'objectivité, le pouvoir discriminatif d'un test est augmenté lorsqu'il se concentre sur chaque item de test individuel et l'améliore. Après avoir donné le test, une analyse d'items peut être réalisée, ce qui montrera la difficulté relative de chaque item et dans quelle mesure chacun fait la distinction entre les bons et les mauvais élèves. Souvent, comme dans le cas de l'obtention de la fidélité, il est nécessaire de prolonger la longueur du test pour avoir une discrimination nette. Un test discriminatif :

- (1) produit une grande variété de résultats lorsqu'il est donné aux élèves qui ont des rendements sensiblement différents
- (2) comportera des items couvrant tous les niveaux de difficulté. Uniquement les meilleurs élèves répondront correctement à certains items; les autres seront relativement faciles et la plupart des élèves y répondront correctement. Si tous les élèves répondent à un item correctement, le test manque de discrimination.



Représentativité

Pour qu'un test soit représentatif, il devrait couvrir les principaux objectifs de cours. Il n'est pas nécessaire ni pratique d'évaluer chaque objectif enseigné en classe, mais un nombre suffisant d'objectifs devraient être inclus pour assurer une mesure valide du rendement de l'élève dans le cours complet.

Validité

La validité est la caractéristique la plus importante d'un bon examen; c'est-à-dire qu'elle indique ce qu'un test est censé mesurer. Il est primordial qu'un test soit valide pour que les résultats soient appliqués et interprétés précisément.

La validité n'est pas déterminée par une seule statistique, mais bien par une recherche qui démontre la relation entre le test et le comportement qu'il veut évaluer. Il y a trois types de validité :

a) Validité du contenu

Lorsqu'un test a la validité du contenu, les items du test représentent l'étendue complète des items possibles que le test devrait couvrir. Les questions du test individuel peuvent être prises d'une grande liste d'items qui englobent une grande variété de sujets.

Dans certains cas où un test évalue un trait difficile à définir, un juge spécialisé peut évaluer la pertinence de chaque item. Deux juges indépendants évaluent le test séparément, car chaque juge fonde son évaluation sur l'opinion. Lorsque les deux juges considèrent que des items sont très pertinents, ils seront inclus dans l'examen final.

b) Validité critérielle

On dit qu'un test a la validité critérielle lorsqu'il est efficace pour prédire les critères ou indiquer une construction. Il existe deux différents types de validité critérielle :

La validité concourante est lorsque les critères sont obtenus en même temps que les résultats. Elle indique dans quelle mesure les résultats estiment avec précision l'état actuel d'une personne pour ce qui est des critères. Par exemple, si un test évaluait les niveaux de dépression, on aurait dit de ce test qu'il avait une validité concourante s'il avait évalué les niveaux de dépression actuels de celui qui subit le test.

La validité prédictive est lorsque les critères sont obtenus après le test. Les tests de carrière ou d'aptitude, qui sont utiles pour déterminer qui est le plus susceptible d'avoir du succès ou non dans certains sujets ou métiers, sont des exemples de tests possédant la validité prédictive.



c) Validité conceptuelle

Un test possède la validité conceptuelle s'il y a une association entre les résultats et la prédiction d'un trait théorique. Les tests d'intelligence sont un exemple d'instruments de mesure qui devraient posséder la validité conceptuelle. L'enseignant peut s'assurer que les items de son test soient valides en suivant la procédure acceptée pour la construction de test qui comprend :

- (1) L'utilisation d'objectifs de cours qui servent de base aux exigences du test. Un examen construit de cette façon aura tendance à évaluer ce qui a été enseigné.
- (2) Une révision des items du test et l'examen complet par les autres enseignants.
- (3) La sélection de la forme la plus adéquate de test et de type d'items de test. Ainsi, si l'enseignant souhaite évaluer « l'aptitude à faire », il doit choisir cette forme de test qui nécessitera que l'élève démontre son « aptitude à faire ». Si une autre forme moins souhaitable est employée, on doit reconnaître que la validité de cette évaluation a été réduite.
- (4) Les exigences du test doivent être présentées clairement et sans équivoque. Si le matériel d'examen ne peut pas être interprété correctement par l'élève, celui-ci ne comprendra pas ce qui est couvert; donc, il ne sera pas apte à répondre comme prévu. Un tel test n'est pas valide.
- (5) L'élimination des facteurs non liés à l'évaluation des points à enseigner, dans la mesure du possible. Un test qui est au-dessus des moyens des élèves par rapport au temps ou au niveau scolaire peut ne pas réussir à évaluer leur apprentissage réel dans le cours.

Fiabilité

La fiabilité se rapporte à la fidélité d'une mesure. On estime qu'un test est fiable lorsqu'on obtient le même résultat à plusieurs reprises. Par exemple, si un test est conçu pour évaluer un trait (comme l'introversion), chaque fois qu'il est donné à un sujet, il devrait donner sensiblement les mêmes résultats. Malheureusement, il est impossible d'évaluer la fiabilité avec exactitude, mais il existe plusieurs façons différentes de l'évaluer.

a) Fiabilité de test-retest

Pour évaluer la fiabilité de test-retest, le test est donné deux fois à deux moments différents. Cette sorte de fiabilité est utilisée pour évaluer la constance d'un test au fil du temps. Ce type de fiabilité suppose qu'il n'y aura pas de changement dans la qualité ou le concept évalué. Le meilleur usage de la fiabilité de test-retest est lorsqu'elle sert pour les choses qui ne changent pas au fil du temps, comme l'intelligence. Généralement, la fiabilité sera meilleure si peu de temps s'écoule entre les tests.



b) Coefficient d'objectivité

Ce type de fiabilité est évalué par deux juges indépendants ou plus qui corrigent le test. Les résultats sont alors comparés entre eux pour déterminer la constance des évaluations des juges. Une manière d'évaluer le coefficient d'objectivité est de demander à chaque juge d'assigner un résultat à chaque item du test. Par exemple, chaque juge pourrait corriger les items en utilisant une échelle allant de 1 à 10. Ensuite, vous pourriez calculer la corrélation entre les deux évaluations pour déterminer le niveau du coefficient d'objectivité. Un autre moyen d'évaluer le coefficient d'objectivité est de demander aux juges de déterminer dans quelle catégorie chaque observation se retrouve et ensuite, de calculer le pourcentage d'accord entre les juges. Donc, si les juges s'entendent 8 fois sur 10, le test a un taux de 80 % de coefficient d'objectivité.

c) Fidélité des formes parallèles

La fidélité des formes parallèles est évaluée en comparant différents tests qui ont été conçus avec le même contenu. Pour y arriver, on produit une grande liste d'items de test qui évaluent la même qualité et après, on sépare les items au hasard en deux tests différents. Les deux tests devraient ensuite être administrés simultanément aux mêmes sujets.

d) Fidélité de la cohérence interne

Ce type de fidélité est utilisé pour estimer la cohérence des résultats pour les items du même test. Essentiellement, on compare des items des tests qui évaluent le même concept afin de déterminer la cohérence interne des tests. Lorsque vous voyez une question qui ressemble beaucoup à une autre question du test, cela peut indiquer que les deux questions sont utilisées pour évaluer la fidélité. Les deux questions sont semblables et élaborées pour évaluer la même chose, alors celui qui subit le test devrait répondre de la même manière aux deux questions, et cela indiquerait que le test possède une cohérence interne.

Les facteurs suivants auront une incidence sur la fidélité d'un test :

- (1) *Administration*. Il est primordial que chaque élève ait le même temps, le même équipement, les mêmes instructions, la même assistance et le même environnement pour effectuer le test. Les instructions du test devraient être appliquées étroitement.
- (2) *Notation*. L'objectivité dans la notation contribue à la fidélité. Chaque effort devrait être déployé pour obtenir l'uniformité des normes et des pratiques de la notation.



- (3) *Normes.* Les normes de rendement établies pour une classe devraient être conformes à celles utilisées dans d'autres classes. Un changement dans les politiques de classement non fondé sur des faits, des normes uniformes et des facteurs d'expériences obtenus des autres classes aura une incidence sur la fidélité des résultats du test.
- (4) *Instruction.* La fiabilité des résultats aux tests sera modifiée si l'instruction présentée à la classe a tendance à trop accorder d'importance aux points à enseigner compris dans l'examen. On appelle souvent ce phénomène « l'enseignement du test » et cela n'est pas souhaitable. Lorsque l'enseignant fournit des indices évidents aux élèves sur les exigences du test, non seulement il modifie la fiabilité du test mais en plus il insulte l'intelligence de sa classe
- (5) *Longueur.* Plus les élèves répondent à des questions, plus le test ou le dispositif de mesure sera fidèle.

Spécification des conditions d'administration

Un bon test doit préciser les conditions dans lesquelles le test sera dirigé. Elles doivent fournir à tous les élèves une bonne chance de démontrer leurs compétences, car cela améliorera la fidélité du test. Le test standardisé doit venir avec la spécification des conditions dans lesquelles les élèves doivent accomplir le test. Donner le test dans des conditions très variées aura une très grande incidence sur les résultats du test. En général, lorsque l'on administre un test, on doit garder à l'esprit les points suivants :

- la disposition physique ;
 - o l'éclairage, la ventilation, le silence, etc. ;
- l'état psychologique ;
 - o éviter de provoquer l'anxiété liée au test ;
 - o essayer de diminuer l'anxiété liée au test ;
 - o ne pas donner le test lorsque d'autres événements seront une source de distraction ;
- les suggestions ;
 - o ne pas parler inutilement avant le test ;
 - o minimiser les interruptions ;
 - o ne pas fournir d'indices aux personnes qui demandent des renseignements sur les items ;
 - o décourager la tricherie ;
 - o allouer aux élèves le même temps pour effectuer le test.

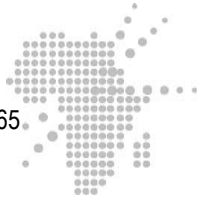


Indications pour la notation et l'interprétation des résultats du test

Un bon test doit être accompagné d'indications qui expliquent comment noter et interpréter les résultats du test, et cela prévaut surtout pour les tests standardisés, qui sont utilisés par les personnes autres que celles qui les ont préparés en premier. De telles indications sont aussi importantes pour les tests conçus localement, car d'autres personnes peuvent prendre part au processus de notation et d'interprétation du test, en raison de circonstances imprévues.

Les directives suivantes permettent d'avoir des indications claires pour les tests.

- 1 Utiliser des descriptions claires pour la procédure détaillée à suivre pour administrer des tests d'une manière standardisée.
- 2 Fournir des renseignements à ceux qui subissent ou utilisent le test en ce qui concerne les formats des questions, la procédure pour répondre aux questions et l'emploi de tout matériel ou équipement requis.
- 3 Établir et instaurer une procédure pour garantir la sécurité du matériel pour le test durant toutes les étapes de l'élaboration, de l'administration, de la notation et du bulletin du test.
- 4 Donner la procédure, le matériel et les indications pour la notation des tests et la surveillance de la justesse du processus de notation. Si le concepteur du test est responsable de la notation, dispenser la formation adéquate aux correcteurs.
- 5 Corrigez les erreurs qui ont une incidence sur l'interprétation des résultats et transmettez immédiatement les résultats corrigés.
- 6 Définir et instaurer une procédure pour s'assurer de la confidentialité des résultats.



Choix d'un test

Généralement, lorsque nous parlons du choix des instruments de mesure ou des tests, nous faisons référence aux instruments standardisés. Dans la plupart des situations, il se peut que vous soyez contraints à une seule option. Toutefois, dans certains cas, vous pourriez être confrontés au défi de comparer des instruments et choisir celui qui est le plus approprié à vos besoins et votre situation. Si tel est le cas, vous devrez baser votre décision sur des faits objectifs et de solides principes d'évaluation.

Avant tout, vous devez penser à la technique d'évaluation qui répond au but du testage et aux objectifs devant être évalués. Par exemple, le plan d'évaluation pour un programme d'études en lecture à l'essai impliquerait une combinaison d'instruments conçus localement et standardisés. Ce plan peut aussi impliquer des formats normatifs et critériels. Par contre, le programme de test de l'école nécessiterait uniquement l'utilisation des tests standardisés.

Lorsque vous devez choisir un test standardisé parmi un certain nombre d'options, vous devez aussi considérer les facteurs suivants pour être sûrs que vos informations soient complètes en ce qui a trait aux caractéristiques du test. Cela peut paraître étonnant, mais il y a un lien entre la quantité d'informations fournies par l'éditeur et la qualité du test. La validité du test est l'information la plus importante requise. De plus, les données concernant la fidélité du test devraient être accessibles et conformes aux normes que nous avons établies pour notre exercice de test. Les autres données qui devraient être transmises comprennent des renseignements sur l'administration, la notation et l'interprétation du test.



Évaluation formative

Répondez aux questions suivantes.

1. Quels types de données d'évaluation sont collectés dans les écoles secondaires de votre région? Quel est le but de chacune de ces données? Selon vous, quelles données sont manquantes et quelle sera l'utilité de telles données?
2. Parmi les types suivants de tests, lesquels sont utilisés dans votre région? À quel niveau chacun est-il utilisé et pour quels objectifs?
 - a. Tests standardisés
 - b. Tests conçus localement
 - c. Tests de rendement
3. Lecture obligatoire pour l'étude #6 et rédiger un court compte rendu (d'environ 500 mots) pour défendre les tests standardisés contre les arguments énoncés dans l'article.
4. Quelle norme de rendement (critérielle ou normative) est utilisée dans votre région? Quels sont ses avantages et ses désavantages?



Activité d'apprentissage #4

CONCEPTION D'INSTRUMENTS

Résumé

Cette activité d'apprentissage est élaborée pour vous aider à comprendre les principes sous-jacents à la conception des instruments de test. L'unité débute par le processus de limitation, de définition et de sélection du contenu qui devrait être couvert par le test. L'utilisation de tableau de spécifications simplifie le processus et permet au concepteur de test de voir le contenu devant être évalué, à quel niveau il faut l'évaluer et la manière de l'évaluer. Ensuite, les indications sur la construction des items du test sont abordées et comprennent le processus de sélection d'objectifs et de questions de compte rendu efficaces. On retrouve des recommandations sommaires pour l'élaboration et le format des tests à la fin de l'activité d'apprentissage. En dernier lieu, on devrait insister sur le fait que les instruments de test doivent être examinés minutieusement avant d'être administrés aux tests.

Liste de concepts-clé

Un tableau de spécifications est essentiellement un projet de test dont le but est de s'assurer que tous les résultats visés, et seulement les résultats visés, sont évalués et que le test contienne un nombre conséquent de questions pour chaque item évalué.

La validation du test est le procédé par lequel un test est administré et revu jusqu'à l'obtention d'un niveau de validité et de fiabilité acceptable.



Contenu

Limitation, définition et sélection du contenu

Lorsque l'on conçoit un test, la première étape est de circonscrire ce qui sera traité. D'ordinaire, les objectifs fournissent le projet pour le sujet ou le groupe de sujets qui sera inclus dans un test, et un test devrait constituer une unité significative de l'enseignement. Normalement, un test couvre l'équivalent de plusieurs jours ou semaines d'enseignement. Limiter la portée du test empêche les élèves d'avoir une surcharge cognitive et permet un échantillonnage beaucoup plus représentatif des comportements représentés par les objectifs choisis. Bien sûr, certains tests cumulatifs comme les examens finaux requièrent une portée plus grande et peuvent représenter des semaines ou des mois d'enseignement. Si les tests individuels ont été conçus rigoureusement, la tâche de la mise au point des tests cumulatifs est facilitée, car les résultats évalués par de tels tests représentent un échantillonnage des rendements évalués par tous les autres tests précédents.

La prochaine étape dans la conception d'un test est d'effectuer un plan détaillé sur le contenu (ou plan de cours), si un tel plan n'est pas accessible. Le plan doit être suffisamment détaillé et comprendre les résultats recherchés des connaissances et des compétences chez les élèves. Bien que tous les résultats ne puissent être évalués dans un test, tous devraient au moins être des candidats pour s'y retrouver. Si des objectifs précis ont été établis, alors les rendements voulus associés à chaque point du plan seront déjà spécifiés. Si ce n'est pas déjà fait, les comportements doivent être identifiés, et nous devons garder à l'esprit que nous nous occupons d'autres aspects que seulement la possession du savoir. Cette manière de décrire le contenu est appelée « description du domaine » et son but est d'identifier tous les items possibles en déterminant les règles pour leur création.

La troisième étape dans l'élaboration d'un test est la conception d'un plan pour échantillonner le domaine de comportement. À l'exception de certains tests critériels, un test représente invariablement un modèle de comportement. Il est pratiquement impossible d'évaluer chaque aspect de chaque résultat. En outre, si nous choisissons minutieusement les aspects que nous évaluons vraiment, nous pouvons généraliser au domaine complet avec un degré de confiance raisonnable. Il est nécessaire que les items du test constituent un échantillon représentatif des comportements voulus. C'est ce que nous avons appelé la validité du contenu. Plus le contenu est bien structuré et défini, plus il est facile d'accomplir les tâches. La sélection aléatoire, qui permet à tout résultat d'avoir une chance indépendante et égale d'être mis dans le test, mène au processus d'atteinte de la validité du contenu dans un test. Vous pouvez également choisir les résultats qui doivent être choisis



pour le test en stratifiant le contenu d'après des critères propres au contenu (ou à d'autres critères logiques).

La conception d'un tableau de spécifications est une approche méthodique d'élaborer un test. Après avoir précisé et défini le contenu du test et les résultats behavioristes, la prochaine étape logique est la construction d'un tableau de spécifications.

Tableau de spécifications

Essentiellement, un tableau de spécifications est un plan pour un test. Comme son nom le laisse supposer, il définit le contenu du test. Le but principal du tableau est de s'assurer que tous les résultats souhaités, et seulement les résultats souhaités, soient évalués et que le test comporte le nombre approprié d'items pour chaque item évalué. Un tableau de spécifications est un tableau à double entrée possédant un axe qui est essentiellement un sommaire du cours et un autre qui indique les comportements souhaités par rapport au contenu.

Sommaire du cours

Les principaux titres et sous-titres du sommaire du cours sont indiqués le long d'un axe du tableau de spécifications. Celui-ci peut comprendre tous les sujets et les sous-sujets ou uniquement ceux spécialement précisés. Plus un axe est détaillé, plus un test a de chances d'inclure des items pour les sujets voulus. Aussi, dans une plus grande mesure, la complexité du sommaire est une fonction de la complexité du domaine d'intérêt évalué.

Résultats behavioristes

Le long de l'axe du tableau, les comportements ou les rendements pour chaque sujet sont indiqués. Cette approche attire l'attention sur des résultats d'ordre supérieur et garantit que le test ne sera pas composé principalement ou entièrement d'items de niveau de connaissance. Le niveau des items dépendra surtout du niveau des élèves et des objectifs établis antérieurement. Toutefois, il faut noter que les professeurs sont plus susceptibles de négliger les items qui évaluent les compétences d'ordre supérieur, comme l'application, la synthèse et l'analyse. Certains tableaux de spécifications peuvent être très précis et indiquer le sous-domaine qui est lié au résultat. Il faut noter qu'il est plus facile d'élaborer un tableau de spécifications pour un test critique plutôt que pour un test normatif, car les résultats souhaités pour le premier ont déjà été examinés en détail et déterminés.



Nombre d'items

L'intersection des titres du sommaire du cours et des catégories des résultats behavioristes forme des cellules dans lesquelles le nombre approprié (ou le pourcentage) d'items est indiqué. Le nombre d'items attribués pour chaque cellule devrait être fondé sur le nombre d'objectifs ou leur influence (ou importance). En fait, cela veut dire que nous devons mûrement penser à « ce que cela prend » pour évaluer un résultat behavioriste précis pour un sujet donné. La conception d'un tableau de spécifications est une activité de planification et idéalement, elle devrait être effectuée durant l'étape de planification de l'évaluation. Sur le plan logique, le meilleur moment pour élaborer la table est après l'établissement des objectifs spécifiques. Si les objectifs sont conçus pour une session complète, une prochaine étape logique est de les séparer en ensembles représentant des unités d'enseignement maîtrisables. Comme pour les objectifs, s'ils sont établis avant l'enseignement, un tableau de spécifications sert de guide pour l'enseignement, car nous savons quels résultats sont souhaités pour chaque sujet et quelle importance il faut accorder à chacun.



Construction des items du test

Vue d'ensemble

Il existe un certain nombre de directives pour la conception d'un certain nombre de différents types d'items. Lorsque les élèves lisent un item, nous voulons que le seul facteur qui ait une incidence sur leur réponse soit la possession ou non du comportement mesuré, c'est-à-dire, l'atteinte de l'objectif. Et pourtant, un item possède de nombreuses caractéristiques qui peuvent empêcher un élève bien informé d'avoir la réponse correcte ou permettre à un élève mal informé d'obtenir une bonne réponse. Des facteurs comme la manière de formuler la question, la nature des options et les consignes pour répondre aux questions peuvent influencer la capacité des élèves à démontrer ou non une réalisation.

En fait, il y a deux types d'item de test, le compte rendu et l'objectif. Les items objectifs comprennent les réponses courtes, les choix multiples, les vrais ou faux et le pairage. Les items de test peuvent aussi être classés en items à réponse construite ou en items de sélection. Les items du compte rendu et des réponses courtes sont des items à réponse construite, car la réponse doit être construite par l'élève. Les items à choix multiple, les items de type vrai ou faux et d'appariement sont appelés les items de sélection, car l'élève choisit une réponse parmi des options. Les types d'item adéquats pour un test donné dépendent essentiellement de la nature des comportements évalués et de l'enseignement. Il n'y a pas de type d'item précis pour certains genres précis de résultats; tout type d'item peut être utilisé pour évaluer tout résultat behavioriste, quel que soit son niveau taxinomique.

Directives générales

Il est généralement recommandé de ne pas concevoir un test à la hâte. Le professeur ou la personne responsable de l'évaluation devrait avoir suffisamment de temps pour planifier, élaborer et réviser les items. Certains tests, comme les postests, doivent être préparés avant l'enseignement. Par contre, les tests d'unité peuvent être conçus après l'enseignement. Il est recommandé de développer des tests en avance et de réviser si besoin est, ou vous pouvez les concevoir au fur et à mesure. Cette approche présente plusieurs avantages, le plus important étant que la tâche difficile qu'est la construction d'un test peut s'étaler sur un espace de temps.

Les items devraient être exactement conformes au résultat behavioriste ou à l'objectif voulu. En d'autres mots, le comportement provoqué par l'item devrait être précisément le comportement indiqué dans l'objectif. Si un objectif indique que l'élève devrait être capable de manipuler un microscope, l'item correspon-



Le concepteur de l'item ne devrait pas lui demander de faire la liste des parties d'un microscope ou de ses utilisations; il devrait plutôt lui demander de le manipuler. Il s'agit de la correspondance objectif-item.

Le caractère suffisant de l'item est étroitement lié au concept de la correspondance objectif-item. Le caractère suffisant de l'item signifie qu'une bonne réponse à un item (ou des items) représente une preuve suffisante que l'élève a démontré le comportement voulu, ou atteint l'objectif. Tout dépendant de la nature de l'objectif, il se peut que cela soit impossible de développer un tel item; plusieurs items peuvent aussi être requis pour être conformes aux directives liées au caractère suffisant. Toutefois, il faut viser ce but.

L'item devrait communiquer clairement le sens voulu du concepteur du test à ceux qui le subissent; il s'agit d'une exigence capitale. Autrement dit, l'item ne devrait posséder aucune ambiguïté, dans la mesure du possible. L'élève devrait interpréter l'item de la même façon que le concepteur du test souhaiterait qu'il soit interprété. L'item devrait aussi être clair pour tout correcteur compétent.

La difficulté de l'item est une caractéristique de l'item qui peut être maniée par le concepteur de test. Les niveaux de difficulté appropriés sont fonction du type de test et des objectifs sur lesquels le test se base. Dans tout cas, ceux qui préparent le test devraient équilibrer la difficulté des items en mettant des items couvrant tous les niveaux du continuum.

La nouveauté de l'item est un concept qui s'applique surtout à l'évaluation des résultats d'ordre supérieur. Elle se rapporte aux conditions où l'élève n'est pas du tout familiarisé avec la situation présente dans l'item, c'est-à-dire que l'élève n'a pas rencontré antérieurement de telle situation durant l'enseignement. Si cette condition n'est pas remplie, un item qui vise l'évaluation d'un résultat d'ordre supérieur peut être en fait un item de connaissance.

Comptes rendus

Les comptes rendus sont surtout utilisés par les professeurs pour évaluer l'élève. Certains tests standardisés comportent des items de compte rendu, et un effort d'évaluation formelle peut de temps à autre se servir de tels items, mais leur utilisation se limite généralement à la classe. Un compte rendu comporte des items qui demandent à l'élève de rédiger des réponses, généralement longues. Le compte rendu est adéquat pour évaluer les résultats qui mettent en jeu les compétences d'ordre supérieur, comme l'application et l'analyse. Il faut noter que la notation des tests objectifs est généralement subjective. Cela signifie qu'il n'existe pas une seule et unique réponse, mais bien qu'il peut exister un grand nombre d'autres options qui répondraient correctement à la question donnée. Moins de questions



peuvent être posées pour une période donnée, car les réponses des comptes rendus sont plus longues que celles des tests objectifs.

Les items objectifs sont plus difficiles à construire, mais plus faciles à corriger; à l'opposé, les comptes rendus sont difficiles à corriger, mais relativement faciles à construire. Les directives générales pour la construction d'item - correspondance objectif-item, caractère suffisant, communication, difficulté et nouveauté – s'appliquent avec une importance particulière sur la communication, car il n'y a pas d'options de réponse pour aider la clarification de la question et il est donc primordial que les options de réponse voulues soient délimitées clairement, dans la mesure du possible. Les élèves devraient recevoir des consignes sur les exigences et la manière de répondre aux questions. De plus, ils devraient recevoir des directives générales au sujet de la longueur, du temps et de la notation.

Comme nous l'avons vu précédemment, la notation des comptes rendus est un processus long et difficile. Bien qu'elle soit potentiellement un processus très subjectif donnant lieu à une faible fiabilité du correcteur, la part de subjectivité peut être considérablement minimisée par une planification et une correction méticuleuses. Nous pouvons objectiver les comptes rendus de façon à ce que nous puissions préciser les réponses et les autres réponses acceptables; alors, tout correcteur bien informé devrait être capable de déterminer si les réponses sont là. Il existe deux approches essentielles pour la correction des comptes rendus : la méthode analytique et la méthode globale.

- a) La méthode analytique comporte l'identification de tous les aspects d'une réponse parfaite et l'assignation d'une valeur en point à chacun.
- b) La méthode globale, aussi appelée la méthode de cotation ou l'approche holistique, a pour résultat une correction plus subjective et moins fiable, mais exige moins de temps. Elle comporte l'identification de tous les aspects d'une réponse parfaite, mais aucune valeur en point n'est assignée à chacun ; à la place, chaque réponse est évaluée en tant que tout et unité totale, et une évaluation globale est effectuée. Des points sont accordés en fonction de l'exhaustivité des réponses.

Comme directive générale pour les comptes rendus, on recommande de corriger chaque réponse sans savoir qui l'a écrite, c'est-à-dire anonymement. Cette façon de faire empêche la partialité de la correction par des facteurs étrangers, comme les sentiments vis-à-vis certains élèves. Aussi, une bonne idée est de lire et corriger les questions du compte rendu deux fois, avant d'accorder les points. L'idéal serait que cela soit effectué par deux personnes différentes.



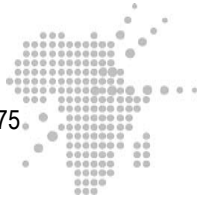
Tests objectifs

Les tests objectifs, dans lesquels les élèves ne doivent pas rédiger des réponses, mais bien choisir parmi un nombre d'options données, sont plus valides et fiables que les items de compte rendu, et la fiabilité du correcteur est beaucoup plus grande. Ces tests peuvent être utilisés pour évaluer des objectifs à tous les niveaux, et ne sont pas limités aux items de niveaux de connaissance, comme plusieurs l'affirment. Les directives générales pour la construction d'item, que nous avons abordées précédemment, s'appliquent généralement. Les recommandations suivantes peuvent vous aider à éviter des pièges communs.

- 1 Faites attention de ne pas donner d'indices sur la bonne réponse.
- 2 Évitez des items dépendants (une bonne réponse pour un item est nécessaire afin de répondre à l'autre item).
- 3 Évitez de complexifier inutilement le test, intentionnellement ou non, par l'emploi d'un vocabulaire complexe ou en rendant l'item plus compliqué pour rien.
- 4 Évitez les tournures négatives et surtout la double négation.
- 5 Évitez les citations directes.
- 6 Les questions insolites ne devraient pas être évaluées.
- 7 Un item devrait avoir seulement une bonne ou une meilleure réponse, sauf si c'est autrement précisé.

Les items objectifs comprennent

- a) les items des réponses courtes
- b) les items à choix multiple
- c) les items de type vrai ou faux
- d) les items d'appariement



Directives générales pour l'organisation, les instructions et le format des tests

Organisation des items des tests

- S'il y a un type d'item, les items du même type devraient être regroupés ensemble.
- Les items peuvent aussi être organisés sur un continuum, du moins difficile au plus difficile.
- Une autre façon logique d'organiser les items est de les classer par sujet. Tous les items connexes au même objectif ou résultat devraient être regroupés ensemble.

Instructions générales

- Si le but et l'importance du test ne sont pas évidents, ils devraient être examinés avec les élèves afin que ceux-ci soient guidés et motivés pour exceller.
- En plus des consignes pour choisir (ou fournir) les réponses, les élèves devraient recevoir des instructions pour noter les réponses.
- Si les instructions sont compliquées ou mettent en jeu une procédure que les élèves ne connaissent pas, un exemple d'item devrait être présenté aux élèves.
- Une bonne idée serait que celui qui subit le test développe le corrigé du correcteur.

Si certains sujets importent plus que d'autres, c'est mieux d'avoir plus d'items pour ces sujets plutôt que d'assigner une plus grande valeur à certains items.

Si le temps requis pour le test est insuffisant, les élèves devraient recevoir des indications pour l'utiliser efficacement.

S'il n'y a pas de pénalité pour les réponses devinées, les élèves devraient être incités à répondre à chaque item, même s'ils sont incertains de la bonne réponse.

- S'il y a une pénalité pour les réponses devinées, on devrait indiquer aux élèves de répondre uniquement aux items pour lesquels ils sont suffisamment certains d'avoir la bonne réponse.



Précisions sur le format

- Les items devraient être numérotés consécutivement et les instructions devraient être placées bien en vue et isolées des items.
- Les espaces pour les réponses devraient être de la même taille et suffisamment grands pour contenir la réponse la plus longue, surtout pour les tests à réponses courtes, et les espaces pour les réponses devraient être placés dans une colonne verticale dans la marge de droite ou de gauche ou sur une feuille de réponses séparée.
- Si un item est accompagné d'une illustration de tout type, comme un diagramme, elle devrait être juste et placée près de l'item, directement au-dessus si possible ou parallèle à l'item.

Révision du test

En fait, la révision du test peut avoir lieu avant son utilisation prévue aussi bien qu'après son administration formelle. Les instruments conçus pour une évaluation précise et formelle peuvent être administrés et révisés encore et encore jusqu'à ce qu'ils atteignent les niveaux acceptables de validité et de fiabilité. Le processus est appelé « test sur le terrain » ou « validation du test ».

Évaluation formative

- 1 Choisissez deux domaines de votre spécialisation ou champs d'intérêt du programme de votre école secondaire. Vous pouvez utiliser les mêmes sujets que lors de l'activité d'apprentissage n° 2.
- 2 Exposez brièvement les sous-sujets des sujets que vous avez choisis. Ils devraient provenir du programme (mais si le programme ne fournit pas de tels renseignements, vous pouvez en développer à l'aide d'un professeur d'école secondaire ou de manuels pertinents).
- 3 La lecture pour l'étude #7 et concevez un tableau de spécifications pour les deux sujets. Un plan d'échantillonnage pour le tableau de spécification approprié est fourni ci-dessous.
- 4 Préparez les questions des tests conformément aux directives précisées dans les lectures #8, #9, #10 et #11.



Activité d'apprentissage #5

ANALYSE, INTERPRÉTATION ET UTILISATION DES RÉSULTATS DE TEST

Résumé

Comme le titre l'indique, cette activité est conçue pour réviser les concepts liés à l'analyse, l'interprétation et l'utilisation des résultats de test. Dans le processus d'analyse, on nous fait remarquer que les statistiques descriptives sont très utiles au professeur pour résumer de façon intelligible les résultats des élèves. Nous avons déjà expliqué qu'un résultat en lui-même est sans importance, à moins qu'il soit interprété à l'aide de critères. Différentes méthodes d'interprétation de résultats de test sont présentées brièvement dans cette activité d'apprentissage. L'unité se termine par une analyse sommaire des différentes façons d'utiliser les résultats de test par différentes parties intéressées dans le processus de l'enseignement.

Liste de concepts-clé

La note brute est le nombre de questions correctes obtenues à un test par l'élève (si l'on suppose que chaque question vaut un point).

Le pourcentage d'exactitude est obtenu lorsque la note brute est divisée par le nombre total de questions et ce résultat est multiplié par 100

L'équivalent de niveau scolaire est un nombre qui décrit la position de l'élève sur un continuum d'accomplissement.

Le rang centile est un résultat indiquant le pourcentage d'élèves dans un groupe précis qui a obtenu des notes brutes plus basses à un test qu'un élève en particulier.



Contenu

Analyse des résultats de test

L'analyse des résultats de test permet au professeur et à tous les utilisateurs d'avoir accès à des données résumées qui peuvent être facilement interprétées au lieu d'avoir de longues feuilles comprenant le nom des élèves et les résultats. La première étape dans l'analyse de données est de décrire ou de résumer les données par l'intermédiaire des statistiques descriptives. Par exemple, il se peut qu'un professeur s'intéresse uniquement à la description du rendement de la classe et du niveau de rendement des personnes par rapport au rendement moyen de la classe. Les statistiques descriptives nous permettent de décrire valablement de nombreux résultats avec peu d'indices en donnant des réponses à des questions comme : quelle a été la note moyenne du test? est-ce que la plupart des notes étaient proches de la moyenne ou considérablement dispersées? et ainsi de suite. Les statistiques descriptives nous permettent également de formuler des énoncés interprétatifs (normatifs et critériels, mais surtout normatifs) sur les scores individuels et de répondre à des questions comme : est-ce que ce score est au-dessus ou en dessous de la moyenne? quel est l'écart de ce score avec la moyenne? quel pourcentage des scores était plus bas que ce résultat? et ainsi de suite. Que des instruments conçus localement ou standardisés soient administrés, les statistiques descriptives constituent un outil important pour interpréter les résultats des résultats.

Voici des types de statistiques descriptives :

1. Évaluation de la tendance centrale
 - a. Le mode
 - b. La médiane
 - c. La moyenne
2. Mesures de variation
 - a. L'étendue
 - b. La déviation quartile
 - c. L'écart type
 - d. La courbe normale
3. Mesures de la position relative
 - a. Rangs centiles
 - b. Notes standardisées
 - c. Équivalent de niveau scolaire et équivalent d'âge
4. Mesures de la relation
 - a. R de Pearson



Interprétation des résultats de test

Trois des buts fondamentaux du test sont (1) de décrire le niveau de développement de chaque élève dans une aire d'expérience, (2) d'identifier les forces et les faiblesses relatives de l'élève par matières et (3) de surveiller le développement d'une année à l'autre en ce qui a trait aux compétences de base. Pour accomplir n'importe lequel de ces buts, il est important de choisir le type de score parmi ceux exposés qui permettra la bonne interprétation. Des scores comme les rangs centiles, les équivalents de niveau scolaire et les notes standardisées diffèrent sur le plan des buts poursuivis, de la précision avec laquelle ils décrivent le rendement et du genre d'informations qu'ils procurent. Une étude plus approfondie de ces types de scores vous aidera à différencier leurs fonctions et les significations qu'ils peuvent exprimer.

Types de scores

Note brute (NB)

La note brute de l'élève est le nombre de questions correctes obtenues à un test par l'élève (si l'on suppose que chaque question vaut un point). Une note brute en soi ne signifie pas grand-chose ou rien. La signification dépend du nombre de questions présentes dans le test et du niveau de difficulté des questions. Par exemple, si Olouch obtient 10 bonnes réponses dans un test de mathématiques et 10 bonnes réponses dans un test en sciences, il serait illogique de conclure que son niveau de rendement dans les deux matières est le même. Cet exemple illustre bien pourquoi les notes brutes sont généralement converties en d'autres types de scores aux fins de l'interprétation.

Pourcentage d'exactitude (PE)

Le pourcentage d'exactitude est obtenu lorsque la note brute est divisée par le nombre total de questions et ce résultat est multiplié par 100. Tout comme les notes brutes, le pourcentage d'exactitude en soi ne veut pas dire grand-chose. Il indique le pourcentage des questions correctes obtenues par l'élève à un test, mais à moins d'en savoir un peu sur la difficulté générale du test, cette information n'est pas très utile. Parfois, le pourcentage d'exactitude est mésinterprété comme un rang centile, qui est décrit plus loin. Les deux sont complètement différents.



Équivalent de niveau scolaire (ENS)

L'équivalent de niveau scolaire est un nombre qui décrit la position de l'élève sur un continuum d'accomplissement. Le continuum est une ligne numérotée qui décrit le niveau le plus bas de connaissances ou de compétences à un bout (les chiffres les plus bas) et le niveau le plus élevé de développement à l'autre bout (les chiffres les plus élevés). L'ENS est un nombre décimal qui décrit le rendement en fonction du niveau scolaire et des mois. Par exemple, si un élève de sixième année obtient un ENS de 8,4 à un test de vocabulaire, il obtiendrait probablement le même résultat qu'un élève typique de huitième année à la fin de son quatrième mois. L'ENS d'une note brute donnée pour tout test indique le niveau scolaire auquel l'élève typique obtient cette note brute. Les chiffres à la gauche du signe décimal représentent l'année et ceux à la droite représentent le mois de l'année.

L'équivalent de niveau scolaire est particulièrement utile et pratique pour évaluer le développement individuel d'une année à l'autre et l'état de développement de l'élève en fonction du niveau scolaire. Cependant, l'ENS a été critiqué, car il est parfois utilisé à mauvais escient et l'on considère « qu'il est facilement mal interprété ». Un problème est de savoir si l'ENS indique le niveau scolaire dans lequel un élève devrait être placé et ce problème est une source de confusion. Par exemple, si un élève de quatrième année obtient un ENS de 6,2 à un test de lecture de quatrième année, devrait-elle être transférée en sixième année ? Il est évident que le niveau de développement de l'élève en lecture est élevé par rapport à ses pairs en quatrième année, mais les résultats du test ne fournissent aucun renseignement sur la façon qu'elle aborderait le matériel lu normalement par les élèves de sixième année, lors des premiers mois. Ainsi, l'ENS n'évalue que le niveau de développement de l'élève ; il ne fait pas de recommandation pour le changement de niveau scolaire. Un ENS beaucoup plus haut ou bas que le niveau scolaire de l'élève est essentiellement un signe de rendement exceptionnel.

En résumé, tout score de test, peu importe le type de score ou le test, est susceptible d'être mal interprété ou utilisé à mauvais escient. Tout score possède des limites ou des faiblesses qui sont exagérées à cause d'une utilisation incorrecte du score. Le secret est de choisir le type de score approprié qui vous permettra le mieux d'accomplir vos buts pour le test. L'équivalent de niveau scolaire convient particulièrement à l'évaluation de l'état de développement de l'élève ou du développement individuel d'une année à l'autre. L'équivalent de niveau scolaire est inadéquat pour trouver le rang de l'élève dans un groupe ou pour identifier des forces et des faiblesses relatives.



Note standardisée du développement (NS)

Comme l'équivalent de niveau scolaire (ENS), la note standardisée du développement est aussi un nombre qui décrit la position de l'élève sur un continuum d'accomplissement. Le principal inconvénient de l'interprétation des notes standardisées du développement est qu'elles n'ont pas de significations inhérentes. À la différence des équivalents de niveau scolaire, qui établissent le niveau scolaire dans le score, les notes standardisées du développement ne sont pas connues de la plupart des pédagogues, des parents et des élèves. Pour interpréter les NS, les valeurs rattachées au rendement typique de chaque année doivent être utilisées comme points de référence.

Le principal avantage de la note standardisée du développement est qu'elle reflète mieux la réalité que l'équivalent de niveau scolaire, c'est-à-dire qu'elle indique que le développement d'une année à l'autre n'est pas normalement aussi bon pour les dernières années que cela l'est pour les premières années. (N'oubliez pas que l'équivalent de niveau scolaire indique le même développement annuel moyen – 10 mois – entre tout couple d'années.) Malgré cet avantage, la note standardisée du développement est beaucoup plus difficile à interpréter que l'équivalent de niveau scolaire. Par conséquent, lorsque les professeurs et les conseillers souhaitent évaluer le développement annuel d'un élève ou le niveau de développement actuel, l'équivalent de niveau scolaire est un choix de prédilection.

Les sources potentielles de confusion et de mauvaise interprétation décrites dans le paragraphe précédent sur les ENS s'appliquent également aux NS. Par rapport à l'ENS, la NS n'est pas aussi facile à utiliser pour décrire le développement, mais elle est aussi inadéquate pour identifier les forces et les faiblesses relatives d'un élève ou pour décrire le rang de l'élève dans un groupe.

Rang centile (RC)

Le rang centile est un résultat indiquant le pourcentage d'élèves dans un groupe précis qui a obtenu des notes brutes plus basses à un test qu'un élève en particulier. Il montre la position ou le rang relatif de l'élève dans un groupe d'élèves qui sont dans la même année et qui ont subi le test au même moment de l'année (automne, mi-année ou printemps) que l'élève. Par exemple, si Toni a obtenu un rang centile de 72 au test en langue, cela veut dire qu'elle a eu un résultat plus haut que 72 pour cent des élèves dans le groupe avec qui elle est comparée. Évidemment, cela signifie aussi que 28 pour cent du groupe a obtenu une meilleure note que Toni. Les rangs centiles vont de 1 à 99.

Le rang centile d'un élève peut varier selon le groupe utilisé pour déterminer le classement. Un élève fait partie simultanément de plusieurs groupes différents : tous les élèves de sa classe, son école, son arrondissement scolaire, son état et le pays.



Types d'interprétation des résultats

Un test de rendement est conçu pour aider à déterminer la quantité de compétences ou de connaissances acquises par l'élève dans une matière précise. Nous utilisons de tels tests pour découvrir si les élèves en savent autant que nous le souhaitons ou s'ils savent des choses précises que nous estimons importantes. En soi, la note brute obtenue à un test de rendement n'indique pas la quantité de connaissances maîtrisées par l'élève ou de compétences acquises. Davantage de renseignements sont nécessaires pour décider de « la quantité ». Le score du test doit être comparé ou faire référence à quelque chose afin d'être significatif. Ce « quelque chose » est d'habitude (a) les résultats obtenus par les autres élèves au test ou (b) une série de descriptions détaillées qui indiquent ce que savent les élèves à chaque point du score ou quelles compétences ils ont démontrées avec succès. Ces deux manières de donner du sens à un score par des références s'appellent communément interprétation normative et interprétation critérielle.

Interprétation normative

Les batteries standardisées de tests de rendement scolaire sont conçues principalement pour donner des interprétations normatives des résultats obtenus pour elles. Pour cette raison, elles sont appelées communément tests normatifs. Toutefois, les résultats permettent aussi des interprétations critérielles, ce que font aussi les résultats de la plupart des autres tests. Ainsi, les tests normatifs sont conçus pour améliorer les interprétations normatives et permettent également des interprétations critérielles.

Une interprétation normative implique la comparaison du résultat de l'élève avec les résultats obtenus au même test par les autres élèves. La quantité de connaissances maîtrisées par l'élève est déterminée par son rang dans le groupe de référence. Un rang élevé signifie que l'élève sait beaucoup de choses ou qu'il est très habile et un rang bas signifie le contraire. Évidemment, les compétences globales du groupe de référence ont une incidence significative sur l'interprétation. Un rang élevé dans un groupe peu habile peut représenter un rendement absolu plus bas qu'un rang bas dans un groupe exceptionnellement très performant.

La plupart des résultats sur les relevés de notes des tests standardisés se fondent sur la normativité, c-à-d la comparaison avec un groupe de référence. Pour ce qui est des rangs centiles, des stanines et des équivalents de la courbe normale, la comparaison s'effectue avec un seul groupe d'élèves dans une année d'études précise qui ont subi l'examen à un certain moment de l'année. Ils sont appelés les résultats de position, car ils indiquent la position ou le rang de l'élève dans un groupe précis. Toutefois, dans le cas des équivalents de niveau scolaire et des notes standardisées du développement, la comparaison se fait avec une série de groupes de référence. Par exemple, les rendements des élèves de troisième année,

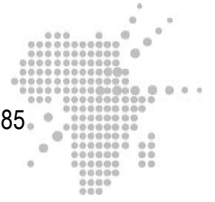


de quatrième année, de cinquième année et de sixième année sont reliés pour former un continuum du développement. (En réalité, l'échelle est composée de groupes d'années d'études allant de l'école maternelle à la fin de l'école secondaire.) Ceux-là sont appelés les résultats de développement, car ils indiquent les positions des élèves sur une échelle de développement. Ainsi, les résultats de position dépendent d'un seul groupe pour dresser des comparaisons et les résultats de développement dépendent de plusieurs groupes qui peuvent être reliés pour constituer une échelle de développement.

Une batterie de tests de rendement scolaire est un ensemble de tests dans plusieurs disciplines et ces tests ont tous été standardisés avec le même groupe d'élèves; c'est-à-dire que les normes pour tous les tests ont été obtenues à partir d'un seul groupe d'élèves pour chaque niveau scolaire. Cet aspect particulier de la batterie de tests de rendement scolaire rend possible l'utilisation de scores pour déterminer les forces et faiblesses relatives par compétence pour les élèves ou les groupes de classe, et l'évaluation du développement d'une année à l'autre. L'utilisation d'une batterie de tests qui possède un groupe de référence commun permet aux pédagogues d'énoncer des affirmations comme « Asha est meilleure en mathématiques qu'en lecture » ou « Ayan est à un niveau plus bas de développement en aptitudes langagières que l'élève typique dans son année d'études ». Si les normes n'étaient pas accessibles, il n'y aurait pas de fondement pour des affirmations comme celles-là.

Les normes permettent aussi la comparaison d'élèves avec d'autres élèves et la comparaison d'écoles avec d'autres écoles. Si la seule raison d'utiliser une batterie standardisée de tests de rendement scolaire était l'établissement de ces comparaisons, alors le temps, les efforts et les coûts rattachés au test devraient être remis en cause. Toutefois, de telles comparaisons donnent l'occasion aux pédagogues de regarder les niveaux de rendement des élèves par rapport à un groupe représentatif d'élèves à l'échelle nationale. Ainsi, les professeurs et les gestionnaires peuvent avoir un regard « externe » sur le rendement de leurs élèves, un regard indépendant des propres évaluations de l'école sur l'apprentissage de l'élève. Tant que notre population continue d'être extrêmement mobile et que les élèves rivalisent à l'échelle nationale plutôt que locale pour des possibilités d'instruction et économiques, les comparaisons avec les élèves, les écoles et un groupe de référence national devraient être d'intérêt pour les élèves, les parents et les pédagogues.

Un malentendu fréquent sur l'utilisation des normes concerne l'effet du test à différents moments de l'année. Par exemple, beaucoup de gens pensent que les élèves effectuant un examen au printemps en quatrième année auront un meilleur résultat que ceux qui font le même examen en automne. Sur le plan des équivalents de niveau scolaire, c'est particulièrement vrai, car les élèves devraient être situés plus haut sur le continuum du développement au printemps qu'en automne.



Par contre, sur le plan des rangs centiles, cette opinion est fautive. Si les élèves de quatrième année ont accompli un progrès typique de l'automne au printemps, leurs rangs parmi les élèves de leur niveau devraient être les mêmes pour les deux saisons. (L'élève ayant un rang centile de 60 en lecture à l'automne est susceptible d'avoir le même rang centile lorsqu'il subira le même test au printemps.) Évidemment, cela s'explique par le fait que différentes normes pour les élèves de quatrième année sont accessibles pour l'automne et le printemps. Bien sûr, les rangs centiles seraient aussi différents que les équivalents de niveau scolaire si les normes pour la quatrième année s'appliquaient pour toute l'année, compte non tenu du moment du test. Ceux qui pensent que les élèves devraient uniquement subir des tests au printemps, car leurs résultats « sonneraient mieux » sont mal renseignés sur la nature des normes et leur rôle dans l'interprétation des résultats.

Les résultats provenant d'un test normatif n'indiquent pas ce que les élèves savent ou ne savent pas; ils montrent seulement comment les connaissances ou les compétences d'un élève donné se comparent avec celles des autres dans le groupe de référence. Uniquement après avoir révisé le plan détaillé sur le contenu du test ou après avoir examiné les items à proprement parler, il est possible d'avancer des interprétations sur ce qu'un élève sait. Toutefois, cette mise en garde ne concerne pas uniquement les interprétations normatives. Pour utiliser un score de test pour déterminer ce qu'un élève sait, nous devons examiner les tâches du test qui lui sont présentées puis déduire ou généraliser à propos de ce qu'il sait.

Interprétation critérielle

Une interprétation critérielle met en jeu la comparaison du résultat d'un élève avec une norme subjective de rendement plutôt qu'avec le rendement d'un groupe de référence. L'interprétation critérielle est nécessaire lorsque l'on prend une décision pour savoir si un élève a maîtrisé une compétence ou a atteint le rendement minimum acceptable. D'ordinaire, les pourcentages d'exactitude sont utilisés et le professeur détermine le score requis pour la maîtrise des notions ou la réussite.

Lorsque vous donnez des interprétations critérielles, il est primordial que la matière couverte par le test-le domaine- soit décrite exhaustivement. Il importe aussi que les questions du test couvrent les aspects centraux du domaine. De plus, il devrait y avoir suffisamment de questions sur le sujet pour donner l'occasion aux élèves de montrer ce qu'ils savent et pour minimiser l'influence des erreurs dans leurs résultats.

Certains tests critériels couvrent une si grande variété de matières ou de compétences qu'il est difficile de donner de bonnes interprétations critérielles à partir des scores du test. Toutefois, dans la plupart des tests, les compétences distinctes sont définies soigneusement et il y a suffisamment de questions qui les évaluent pour que l'on puisse donner concrètement de bonnes interprétations critérielles à partir des scores.



Le pourcentage d'exactitude est le genre le plus largement utilisé pour **donner des** interprétations critérielles. Les scores critères qui déterminent divers niveaux de rendement des tests sont en général des pourcentages d'exactitude obtenus grâce à l'analyse et au jugement du professeur

Utilisation des résultats du test

L'évaluation est la collecte et l'analyse systématique de données à des fins décisionnelles. Ainsi, les résultats de l'analyse des données sont utilisés pour la prise de décisions. Les types de décisions qui viennent après l'évaluation sont extrêmement variés et voici deux exemples : on peut décider qu'un élève X n'accomplit pas de progrès satisfaisant durant cette étape scolaire et conclure que le système éducatif du pays Y n'atteint pas ses objectifs. Les types de personnes suivants, à l'intérieur et à l'extérieur du système scolaire, s'intéressent aux résultats des tests :

- a) Les législateurs
- b) Les commissions scolaires
- c) Les gestionnaires
- d) Les parents
- e) Les élèves
- f) Les décideurs



Évaluation formative

Dans le tableau ci-dessous, les données pour les résultats des examens de mathématiques et d'anglais sont fournies. Ces données ont été collectées à la fin de l'année scolaire. Les élèves qui ont subi les examens étaient en deuxième classe à leur école secondaire.

Effectuez les activités suivantes en utilisant les données fournies.

1. La lecture pour l'étude #13 et la compréhension des différentes méthodes d'analyse des résultats d'examen et ensuite, l'accomplissement des tâches suivantes.
 - a. Tracez un diagramme à barres pour illustrer la répartition des élèves.
 - b. Calculez la moyenne, la médiane et le mode des données.
 - c. Calculez l'écart type des données.
 - d. Calculez la corrélation entre les résultats des élèves obtenus aux deux examens.
2. Interprétez les données en utilisant les deux principales manières d'interpréter les résultats d'examen.
 - a. Utilisez d'abord l'approche normative
 - i. Tracez la courbe normale pour illustrer la répartition des élèves.
 - ii. Combien d'élèves se retrouvent dans chaque catégorie de la courbe?
 - iii. Que pouvez-vous dire à propos du résultat de l'élève # 19? Comment a été le rendement de cet élève?
 - iv. Comment peut-on utiliser ces résultats?
 - v. Quelles sont les faiblesses de cette méthode d'interprétation des résultats de test?
 - b. Deuxièmement, utilisez l'approche critérielle pour interpréter les mêmes données.
 - i. Établissez des critères hypothétiques pour interpréter les données
 - ii. Tracez un histogramme pour illustrer le nombre d'élèves se retrouvant dans chaque catégorie de vos critères.
 - iii. Que pouvez-vous dire à propos du résultat de l'élève # 19? Comment a été le rendement de cet élève?
 - iv. Comment peut-on utiliser ces résultats?
 - v. Quelles sont les faiblesses de cette méthode d'interprétation des résultats de test?



3. La lecture pour l'étude #12 et déterminez comment les parties intéressées suivantes en éducation peuvent utiliser de tels résultats.
- a. Professeur
 - b. Élève
 - c. Parent
 - d. Administrateur scolaire
 - e. Conseiller scolaire
 - f. Ministère de l'Éducation

Université Amoud

Faculté d'éducation

Notes des élèves

Classe: première année

Identification de l'élève	Mathématiques	Anglais I
1.	62	56
2.	45	57
3.	68	78
4.	90	8
5.	89	56
6.	67	88
7.	55	99
8.	44	90
9.	68	98
10.	98	80
11.	78	60
12.	79	40
13.	20	56



14.	49	78
15.	33	88
16.	88	56
17.	99	88
18.	100	77
19.	34	66
20.	56	55
21.	100	23
22.	77	67
23.	78	88
24.	65	34
25.	54	100
26.	21	90
27.	23	99
28.	47	100
29.	67	87
30.	68	77

La note globale pour les deux examens est de 100.

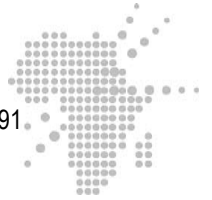


XIII. Évaluation sommative

Cette activité souhaite évaluer ce que vous avez été en mesure d'apprendre dans ce module. Il est temps que nous mettions à profit les notions apprises jusqu'ici sur l'évaluation et les tests en pédagogie pour les appliquer. Donc, en collaboration avec vos professeurs et votre établissement d'enseignement, vous devriez avoir accès à l'enseignement (on recommande l'école secondaire, mais si c'est impossible, le niveau primaire ou l'établissement collégial conviendra). Avant que vous commenciez l'enseignement en soi, vous devriez obtenir les programmes de cours et les manuels nécessaires. Aussi, avant le début de votre enseignement, vous devriez préparer la méthode d'évaluation et l'évaluation que vous mènerez après avoir terminé l'enseignement. Après avoir terminé l'enseignement et dirigé l'examen, vous devez écrire un rapport de 1000 à 2000 mots.

Contenu du rapport

- 1 Page couverture
- 2 Résumé
- 3 Table des matières
- 4 Contexte
5. Étape de la planification
 - a. Analyse de situation
 - b. Spécification des objectifs
 - c. Spécification des préalables
 - d. Choix et élaboration des instruments de mesure
 - e. Présentation des stratégies
 - f. Préparation d'un calendrier général
6. Étape du processus
 - a. Prétest
 - b. Déroulement de l'évaluation
7. Étape du produit
 - a. Résultats du test
 - b. Analyse des résultats
 - c. Interprétation des résultats
 - d. Recommendations aux parties intéressées
 - e. Auto-évaluation
8. Conclusion



Modèle d'évaluation

Les notes globales pour le rapport sont de 100%. Les directives suivantes aideront le professeur du module à noter le rapport et donnent une idée de ce qui est attendu.

1. Page couverture (5 points)
2. Résumé (5 points)
3. Table des matières (5 points)
4. Contexte (5 points)
5. Étape de la planification (30 points)
 - a. Analyse de situation
 - b. Spécification des objectifs
 - c. Spécification des préalables
 - d. Choix et élaboration des instruments de mesure
 - e. Présentation des stratégies
 - f. Préparation d'un calendrier général
6. Étape du processus (15 points)
 - a. Prétest
 - b. Déroulement de l'évaluation
7. Étape du produit (25 points)
 - a. Résultats du test
 - b. Analyse des résultats
 - c. Interprétation des résultats
 - d. Recommandations aux parties intéressées
 - e. Auto-évaluation
8. Conclusion (10 points)



XIV. Synthèse du module

Ce module a couvert la plupart des concepts importants sous-jacents à l'évaluation et aux tests en pédagogie. À la fin du module, on s'attendait à ce que les élèves qui maîtrisaient le contenu abordé dans ce module soient en mesure de comprendre le processus du déroulement d'une évaluation efficace en éducation et d'en diriger une. Ce module possède cinq unités. L'unité 1 (l'évaluation en pédagogie) présente le concept, la nature, les types et les étapes de l'évaluation. Cette unité sert d'introduction à ce module, car il donne aux apprenants une idée de l'évaluation et de son rôle en éducation. La deuxième unité (spécification des objectifs d'éducation) présente en détail l'une des étapes les plus importantes dans le processus de l'évaluation. L'évaluation commence par des objectifs et vise à vérifier le niveau d'atteinte de ces objectifs. Dans cette unité, on enseigne aux apprenants comment établir des objectifs d'éducation mesurables qui représentent toute la gamme des domaines d'apprentissage. Dans la troisième unité, le processus de classification des tests et la sélection de ceux qui répondent aux besoins du professeur sont abordés. Dans cette unité, l'apprenant a l'occasion de voir comment il peut profiter des tests standardisés actuels dans ses tests. La quatrième unité est consacrée à l'élaboration des instruments qui doivent être utilisés dans le processus d'évaluation. Diverses techniques de conception de test, de construction d'items et leur révision sont présentées à l'apprenant. La cinquième et dernière unité se concentre sur les manières d'analyser, d'interpréter et d'utiliser les résultats de test. Cette unité veut aider l'apprenant à penser à l'utilisation des résultats de test et lui donner les aptitudes qui l'aident à se servir des résultats de façon appropriée.



XV. Auteur du module

M. Ridwan Mohamed Osman est actuellement le doyen de la faculté d'éducation à l'Université d'Amoud, au Somaliland. Il est aussi le coordonnateur du programme AVU de formation des enseignants à l'Université. Il a été un chargé de cours en méthodologie de recherche et en éducation à la même université et pour d'autres établissements d'enseignement au cours des cinq dernières années. M. Osman a obtenu son baccalauréat en éducation à l'Université d'Amoud. Il a également obtenu sa maîtrise en éducation à l'Université d'Egerton au Kenya. Ses recherches actuelles portent sur l'enseignement de la science dans les écoles secondaires africaines, la gestion de classe, l'évaluation et les tests en pédagogie et l'éducation des indigènes. Pour de plus amples informations, vous pouvez envoyer un courriel à l'adresse électronique suivante :

ridwaanxaaji@hotmail.com

Tél: +25224457020