

**Validité de quelques épreuves du premier semestre à la Faculté de  
Psychologie et des Sciences de l'Éducation de l'Université de  
Kisangani pour l'année Académique 2016-2017 (Cas du  
Département des Sciences de l'Éducation)**

**Validity of some tests of the first semester at the Faculty of  
Psychology and Educational Sciences of the University of  
Kisangani for the 2016-2017 academic year (case of the  
Department of Educational Sciences)**

**Jean-Claude KAPARATUM PITUWA**  
Enseignant Chercheur en Sciences de l'Éducation  
Université de Bunia  
Province de l'Ituri  
République Démocratique du Congo  
**jckaparatum@gmail.com**

**Date de soumission :** 09/03/2022

**Date d'acceptation :** 08/06/2022

**Pour citer cet article :**

KAPARATUM PITUWA J.C. (2022) «Validité de quelques épreuves du premier semestre à la Faculté de Psychologie et des Sciences de l'Éducation de l'Université de Kisangani pour l'année Académique 2016-2017 (Cas du Département des Sciences de l'Éducation)», Revue Internationale du Chercheur «Volume 3 : Numéro 2» pp : 902 - 917

## Résumé

Cette recherche sur la validité de quelques épreuves à la Faculté de Psychologie et des Sciences de l'Éducation de l'Université de Kisangani avait pour but de vérifier si les examens du premier semestre au Département des Sciences de l'Éducation ont été valides. Concrètement, nous cherchions, d'une part, à voir à la fois combien de questions sont appariées aux objectifs et combien d'objectifs pédagogiques ne sont pas mesurés par les questions des épreuves ; et, d'autre part, à dégager le degré de relation le plus prépondérant entre les questions et les objectifs pédagogiques. Elle a été menée par la méthode transversale appuyée par la technique documentaire. Les résultats obtenus nous ont montré que les épreuves du premier semestre au Département des Sciences de l'Éducation ne mesurent pas globalement les objectifs opérationnels formulés lors des enseignements, car moins de 50 % de questions (35,7 % < 50 %) sont appariées aux objectifs et la plupart d'objectifs (62,2 % > 50 %) ne sont pas mesurés par les questions. En outre, par rapport au degré de relation Q/O, la congruence nulle est le degré le plus prépondérant de rapport entre les questions et les objectifs pédagogiques. En somme, il est de toute urgence de procéder à la formation des enseignants dans la science de construction des questions.

**Mots clés :** Validité ; épreuve ; premier semestre ; Faculté de Psychologie et des Sciences de l'Éducation ; Département des Sciences de l'Éducation.

## Abstract

This research aimed at verifying whether exams of first semester at the Department of education have been valid. The researcher tried to see verify at the same time how many questions were relevant to the objective and how many pedagogical objectives were not measured by different questions of testing and then to get the degree of relations between questions and objectives. Diagonal or transversal method and documentary technique were used for collecting data. The results have shown that first semester's testing at the department of sciences of education did not globally measure the objectives during the teachings because less than 50 % of questions (35.7 % < 50 %) were relevant to the objectives and most of objectives (62.2 % > 50 %) were not measured by questions. Finally, there is an emergency of proceeding to the formation of teachers in the science of constructing good questions.

**Key words:** Validity; testing; first semester; Department of Psychology and sciences of Education.

## Introduction

La validité est le critère le plus important dans le choix d'une épreuve. Elle fait référence aux caractéristiques mesurées par l'épreuve et à l'exactitude avec laquelle elle évalue ces caractéristiques. C'est dans ce contexte que ANGOFF (1988, p.83) rappelle que si le concept de validité a changé, l'importance que lui accordent les psychométriciens est restée constante : « en psychométrie, la validité a toujours été considérée comme le concept le plus fondamental et le plus important ».

Il se révèle qu'un instrument invalide pousse par conséquent à sa désorientation. Il en est de même des épreuves d'acquis scolaires. L'invalidité amène à mesurer d'autres caractéristiques que celles voulues par l'évaluateur. C'est ainsi que, EBEL (1976, p.411) confirme que la validité d'un test est la proportion de la variance vraie et pertinente dans la variance totale observée par rapport aux traits mesurés par le test.

Beaucoup d'études, comme celle de MASANDI (2007, p.41), nous ont montré que la pratique de l'évaluation pose d'énormes difficultés aux enseignants dans leurs activités quotidiennes. Ces derniers se posent les questions du « comment évaluer ? » « Quels instruments utiliser ? ». Plusieurs études expérimentales (LAUGIER, 1934 et 1936 ; NOIZET et al, 1961, 1969, 1975, 1977 et 1978 ; PIERON et al, 1962), ... ont conclu à l'insuffisance des méthodes d'évaluation couramment utilisées dans le domaine de l'enseignement-apprentissage. Ceci montre que l'évaluation reste de nos jours un problème qui nécessite des solutions urgentes étant donné qu'un dysfonctionnement persiste dans son rôle et ses procédures, malgré la diversité d'approches actuellement disponibles.

En outre, il est essentiel de souligner aussi qu'un instrument psychométrique ou éducatif n'est jamais valide dans l'absolu, car MUGENO (2016, p.302) révèle que la validité est toujours établie en rapport à des objectifs, à des populations cibles, ainsi qu'à des contextes d'application spécifiques. En conséquence, il ne s'agit pas de déterminer si un instrument est valide ou pas ; on évaluera plutôt quel est son degré de validité donné pour un objectif précis (par exemple, mesurer l'intelligence), pour une population cible précise (par exemple, les enfants de l'école élémentaire) et pour un contexte d'application spécifique (par exemple, prédire la performance scolaire).

Pour arriver à cette vérification, tout enseignant, dans le cas d'une validité de contenu, par exemple, comme nous le demandent CROCKER et ALGINA (1985, p.213), devrait s'en tenir à la prise en compte de certains indicateurs permettant d'évaluer combien les questions

posées à l'examen représentaient les objectifs opérationnels formulés lors des enseignements. Concrètement, il devrait évaluer le taux des questions associées aux objectifs opérationnels et voire aussi aux objectifs opérationnels jugés très importants. En plus, il devrait veiller non seulement au pourcentage des objectifs non mesurés par les questions, mais aussi à la corrélation entre le poids des objectifs et le nombre de questions qui les mesurent. Enfin, un calcul d'indice de congruence entre les questions et les objectifs devrait le guider dans cette évaluation. Agir ainsi, garantirait la validité du contenu de l'épreuve.

Pourtant, selon KAMBA (2004, pp.41-57), les évaluations auxquelles les enseignants recourent le plus dans leurs classes dérogent à cette règle. Ce syndrome a pris de l'ampleur dans notre système éducatif, en général, et dans nos institutions supérieures, en particulier, et plus spécifiquement à la Faculté de Psychologie et des Sciences de l'Education de l'Université de Kisangani et dans le Département des Sciences de l'Education.

C'est pour cette raison, cogitant sur cette problématique au cours de l'année académique 2016-2017 au premier semestre, nous nous sommes demandé si les examens du premier semestre au Département des Sciences de l'Education de l'Université de Kisangani ont été valides. Explicitement, pour l'ensemble des épreuves en étude, combien de questions sont associées aux objectifs ? Combien d'objectifs ne sont pas mesurés par les questions ? Quel est le degré de rapport prédominant entre les questions et les objectifs pédagogiques ?

Compte tenu du fait que plusieurs enseignants dérogent à la règle de l'évaluation dans leurs classes (KAMBA, 2004, pp.41-57), nous avons pensé que les épreuves du premier semestre au Département des Sciences de l'Education ne mesureraient pas globalement les objectifs opérationnels formulés lors des enseignements. Autrement dit, la majorité de questions (> 50%) ne s'apparieraient aux objectifs, d'une part, et la plupart d'objectifs (> 50 %) ne se mesureraient par les questions, d'autre part. En outre, par rapport au degré de relation Question/Objectif, la congruence nulle serait le degré le plus prépondérant de rapport entre les questions et les objectifs pédagogiques.

Cette étude vise à vérifier la validité de contenu des questions des examens du premier semestre de l'année académique 2016-2017. L'intérêt de cette étude est de réveiller l'attention des chercheurs en matière de l'évaluation des acquis scolaires, leur montrer les problèmes auxquels est confrontée l'évaluation des acquis dans notre système éducatif et sensibiliser les praticiens (les enseignants) à veiller sur les règles conduisant à bien évaluer.

Il sied de signaler qu'outre l'introduction et la conclusion du présent article, il développe les points suivants : le premier porte sur la revue de la littérature, le second point est consacré à la méthodologie de la recherche suivie pour étudier le phénomène. DEPELTEAU (2003), définit la méthodologie comme étant l'ensemble des règles, étapes et procédures auxquelles on a recours dans la science pour saisir les objets étudiés. Le troisième point présente les résultats issus des analyses et enfin dernier développe la discussion des résultats ; concrètement, il a été question de la confrontation des résultats de cette étude à celles des études antérieures.

### **1. Revue de la littérature**

Newton (2012) cité dans André et al., (2015) a identifié les premières traces de la notion de validité dans la documentation de l'année 1915 : ainsi, Terman et ses collègues ont discuté de la validité d'un test d'intelligence et de la validité du QI. Il attribue en outre la première définition des termes de validité et de fidélité à Buckingham et al. (1921), en ces termes : deux des plus importants types de problèmes en mesure sont ceux qui ont trait à la détermination de ce que le test mesure, et à la consistance de cette mesure. Le premier pourrait être appelé le problème de validité et le second le problème de fidélité (Buckingham et al., 1921, p. 80). Au sens originel, le problème de validité est de savoir si un test mesure réellement ce qu'il est censé mesurer (Kelley, 1927), selon Angoff (1988), cette définition prévaut jusque dans les années 1950. Ainsi, à cette époque et pendant plusieurs décennies, la validité est définie comme le degré auquel un test mesure ce qu'il prétend mesurer.

Au fil du XXe siècle apparaissent plusieurs formes de validité, largement utilisées et discutées, et qui ont évolué au fil du temps. Apparition et évolution correspondent d'ailleurs souvent à l'émergence ou au raffinement de techniques statistiques.

Même s'il est question dès 1940 de validité manifeste, qui est une forme dite naïve de validité, c'est la validité en référence à un critère, concomitante ou prédictive, qui prédomine jusque dans les années 1950. Pour Guilford et ses contemporains, ces deux formes de validité en référence à un critère correspondent à un argument sur la valeur utile du testing. Elles représentent une relation de cause à effet entre le score du test et une caractéristique mesurée à un autre moment, celui-ci pouvant être futur ou non.

L'exemple qui suit offre une illustration en éducation et met en évidence une limite. L'habileté de l'élève que l'on cherche à estimer est souvent définie à partir d'un échantillonnage des performances attendues. Ainsi, dans le cas d'un test pour évaluer

l'habileté à résoudre des problèmes en algèbre, le critère peut être fourni par une autre version du test, et la corrélation entre les deux versions peut offrir un argument de validité concomitante. Toutefois, la qualité de cette autre version n'est pas assurée, ce qui remet en question l'argument de validité ainsi obtenu (Kane, 2013).

Schmidt, (2012), rappelle qu'à partir de 1950, la validité manifeste s'affine en validité de contenu, couvrant les facettes de l'univers à quantifier. Il semble opportun de noter que la validité manifeste, pourtant vue comme étant superficielle dans une vision psychométrique, est encore utilisée aujourd'hui pour valider divers instruments de collecte de données, notamment en éducation et pour la sélection de personnel.

Cependant, vers la fin des années 1970 une autre vision de la validité reposant sur un ensemble de modèles et de méthodes a émergé (Kane, 2013). Cette tendance se concrétise avec la validité unifiée de Messick (1989), qui inclut des considérations plus morales en lien avec les conséquences de l'usage du test et de l'interprétation des données. Ainsi, pour lui, un bon argument de validité intègre en un tout cohérent six éléments de preuve de la validité conceptuelle (ou de construit) : le contenu du test, les processus de réponse, la structure interne, les relations avec d'autres variables, la généralisation de la validité et les conséquences du testing.

André et al., (2015) rappellent que des nouvelles visions continuent à émerger. Leur objectif est, dans certains cas, de rapporter la validité à des propriétés plus psychométriques ; dans d'autres, de revenir au concept de validité tel qu'il a initialement été défini et de le rendre moins généraliste ; et, enfin, d'élargir le concept. Les définitions, les formes de validité et les manières de les opérationnaliser continuent donc à faire réfléchir autant les praticiens que les chercheurs et à susciter des débats.

## **2. Modèle conceptuel**

Il sied de rappeler que la validation du contenu d'un test psychométrique diffère tant soit peu de l'évaluation des acquis scolaires. La construction d'une épreuve scolaire est souvent fonction de l'objectif qui lui est assigné car dans l'enseignement les épreuves sont appelées à jouer plusieurs rôles : dresser un bilan des acquis des enseignés, prendre une décision sur la promotion de l'élève, sélectionner les élèves selon certaines caractéristiques particulières afin de former des groupes, préparer une révision de la matière à partir des points pour lesquels certains apprenants éprouvent des difficultés, etc ; par contre, Haynes et al. (1995) nous présentent 7 principes de base pour valider un test psychométrique :

1. Définir avec soins le domaine et les facettes du concept et valider cette définition ;
2. Utiliser un échantillon d'experts et de membres de la population de référence pour créer les items et les autres aspects du test ;
3. Soumettre tous les aspects du test à une validation de contenu ;
4. Utiliser plusieurs experts pour valider le contenu d'un test et quantifier leurs jugements à l'aide d'échelles formalisées ;
5. Examiner la représentation proportionnelle des items relativement aux différentes facettes du concept ;
6. Présenter les résultats de la validation de contenu lors de la publication de tout nouvel instrument ;
7. Présenter en compte toutes les analyses psychométriques ultérieures pour affiner la validation du contenu du test

Il est important de souligner que cette démarche est plus psychométrique qu'éducatrice, car elle concerne essentiellement un test psychométrique et moins une épreuve scolaire. Si nous l'avons évoqué ici, c'est justement pour élucider les opérations censées avoir été faites par les professeurs lors de l'élaboration de leurs épreuves (par exemple le principe 1) et celles que nous avons suivies dans cette étude (par exemple les principes 3,4 et 6).

Etant donné que la démarche de Haynes et al. (1995) n'est pas de manière satisfaisante appropriée au domaine de l'éducation, nous avons eu recours aux indicateurs de validation d'une épreuve telle que décrite par Crocker et Algina (1985) dans le point ci-dessous afin d'atteindre notre objectif dans cette étude.

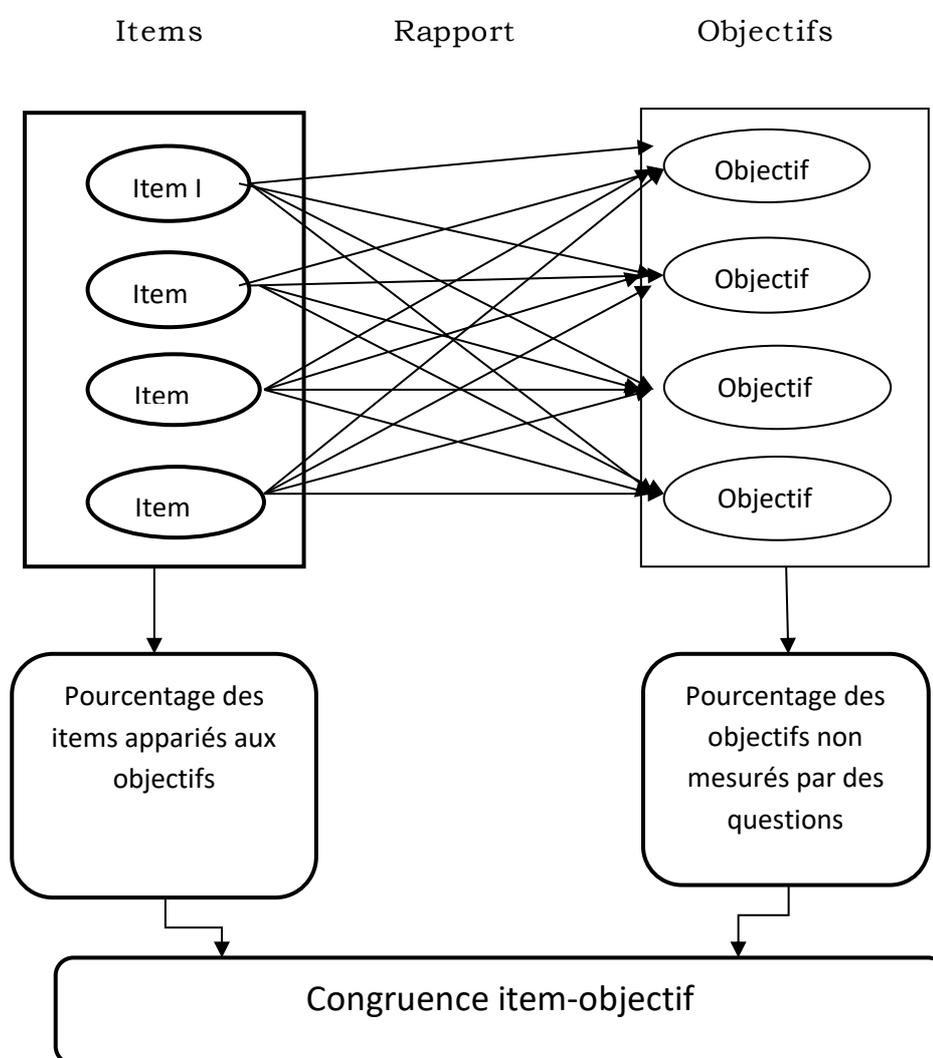
Dans le domaine de l'éducation, Crocker et Algina (1985) énumèrent 5 indicateurs utilisés pour évaluer dans quelle mesure un ensemble d'items est représentatif des objectifs pédagogiques visés par l'épreuve. Il s'agit d'indicateurs suivants : le pourcentage d'items appariés aux objectifs ; le pourcentage d'items appariés aux objectifs jugés très importants ; la corrélation entre le poids des objectifs et le nombre d'items les mesurant ; l'indice de congruence item-objectif et le pourcentage des objectifs non mesurés par les items.

Il convient de noter que les 5 catégories d'indices ne fournissent pas une information équivalente sur la congruence item-objectif. Les deux premiers, en particulier, nécessitent un important échantillon d'items. Enfin, le troisième ne fournit pas de résultats importants si

tous les objectifs sont d'égale importance. Une faible variation des valeurs de pondération de chaque objectif entrainera une diminution de la valeur maximale de la corrélation.

Etant donné la nature de nos données ainsi que la manière dont elles ont été dépouillées, nous avons retenu seulement trois indicateurs, à savoir le pourcentage des questions appariées aux objectifs, le pourcentage des objectifs non mesurés par les questions et l'indice de congruence question/objectif fourni par la formule de calcul de l'indice de Hambleton. Ce sont ces trois indicateurs qui nous ont paru pertinents. N'ayant pas de nous-mêmes définis les objectifs opérationnels, il nous a été très difficile de juger de leur importance, les uns par rapport aux autres. C'est pour cette raison que nous n'avons pas pris en compte les indicateurs 2 et 3 dans le processus de validation de nos épreuves.

**Figure 1 : Modèle conceptuel pour valider une épreuve scolaire**



Source : Auteur

### 3. Matériel et méthodes

Cette étude a porté sur 9 cours du Département de Sciences de l'Éducation de la Faculté de Psychologie et des Sciences de l'Éducation de l'Université de Kisangani pour le compte de l'année académique 2016-2017. Le choix de ces 9 cours du Département des Sciences de l'Éducation est justifié par le fait que non seulement, les enseignants dudit Département sont sensés maîtriser les principes et règles d'élaboration d'une épreuve, mais aussi, ce sont eux qui enseignent tous les cours qui traitent des sciences de construction des questions.

La Statistique descriptive, l'Éducation à la citoyenneté, la Pédagogie générale, la Didactique générale, l'Étude des programmes de l'enseignement primaire et secondaire, la Docimologie, la Pédagogie comparée, l'Apprentissage scolaire, l'Évaluation des apprentissages du domaine affectif et psychomoteur sont les cours qui ont fait l'objet de cette étude.

La technique documentaire nous a été utile pour la récolte des données. Celle-ci nous a permis de rassembler les questions et les objectifs opérationnels de différents cours retenus faisant l'objet d'examens du premier semestre à la Faculté de Psychologie et des Sciences de l'Éducation. Nous avons d'abord confronté les verbes d'action contenus dans les objectifs spécifiques aux verbes utilisés dans les questions posées dans le but d'étudier le degré de leur correspondance.

Il sied de rappeler qu'il est question pour cette recherche de la confrontation des verbes d'action contenus dans les objectifs opérationnels aux questions des examens, parce que c'est supposé que les contenus des différents cours ont été élaborés par rapport aux objectifs opérationnels et que ceux-ci sont le reflet des contenus parce que les objectifs sont censés couvrir l'ensemble de la matière. Ainsi, la validité de contenu compris dans cette optique se résume à la confrontation des questions des examens aux objectifs opérationnels contenus dans les cours. Cette démarche est éclairée par l'idée de CROCKER et ALGINA (1985) qui confirment que lorsqu'on cherche à construire un instrument avec l'objectif de dresser un bilan des apprentissages, un échantillon représentatif et exhaustif du contenu est nécessaire ; pour ce faire, une mesure fondée sur les objectifs d'apprentissage est capitale.

Nous avons ensuite fait appel aux trois juges pour l'appréciation de ces correspondances ; concrètement, ils se sont prononcés sur les relations items-objectifs ; c'est-à-dire répondre à la question « les items sont-ils construits à partir des objectifs opérationnels ? ».

Après confrontation des verbes d'action contenus dans les objectifs opérationnels et les questions, appréciation des relations items-objectifs par les juges, nous avons d'abord trouvé le pourcentage de questions appariées aux objectifs en appliquant la formule suivante :

$$\text{Questions appariées} = \frac{\text{Nombre de questions appariées}}{\text{Nombre de questions}} \times 100$$

Ensuite, le pourcentage d'objectifs non mesurés par les questions a été trouvé par la formule suivante :

$$\text{Pourcentage des objectifs non mesurés} = \frac{\text{Nombre d'objectifs non mesurés}}{\text{Nombre d'objectifs}} \times 100$$

En outre, les moyennes des appréciations de la congruence Question par Objectif par les juges ont été obtenues grâce à la formule qui suit :

$$\text{Moyenne des appréciations de la congruence par des juges} = \frac{\text{Somme des moyennes par objectifs}}{\text{Nombre de juges}}$$

Enfin, l'indice de congruence item-objectif de HAMBLETON (1980) qui a été proposée en version simplifiée par CROCKER et ALGINA (1985, p.365) nous a été utile dans le calcul de congruence ou d'équivalence entre les questions et les objectifs pédagogiques à partir des appréciations des juges sur des échelles. Ces échelles varient de -1 à +1 : -1 indique que la question ne mesure pas l'objectif, 0 montre que le juge est incertain et +1 traduit que le l'item mesure l'objectif. Voici la formule :

$$\text{Indice de congruence Item-Objectif} = \frac{\text{Nombre d'objectifs}}{2 \times \text{Nombre d'objectifs} - 2} (\text{Moyenne des évaluations de l'item } i \text{ pour l'objectif } k - \text{Nombre d'objectifs})$$

Après les analyses par l'indice de congruence item-objectif de HAMBLETON, nous avons abouti aux résultats ci-après, consignés dans un tableau synoptique qui reprend, le pourcentage de questions appariées aux objectifs, le pourcentage des objectifs non mesurés par les questions et l'indice de congruence question-objectif.

## 4. Résultats

Tableau 1. Résultats Synoptiques de la congruence Question par Objectif de toutes les épreuves

N°	Epreuves	Questions appariées		Objectifs non mesurés		Congruence parfait		Congruence satisfaisante		Congruence nulle	
		N	%	N	%	N	Valeur	N	Valeurs < 0.50	N	Valeur=0
1	Statistique descriptive	11	92	3	37	3	1	5	0.51, 0.66, 0.83, 0.90, 0.90	16	0
2	Education à la nouvelle citoyenneté	2	40	3	75	1	1	0	-	8	0
3	Pédagogie générale	4	33	4	50	2	1	4	0.50, 0.66, 0.83, 0.83	40	0
4	Didactique de l'enseignement de Psy-Péda.	2	22	3	60	1	1	1	0.66	35	0
5	Etude de programme de l'enseignement Prim. et Sec.	1	17	5	83	2	1	0	-	24	0
6	Docimologie	1	17	4	83	0	1	1	0.59	24	0
7	Pédagogie compare	0	0	4	100	0	1	0	-	36	0
8	Apprentissage scolaire	2	33	1	33	0	1	0	-	9	0
9	Evaluation des apprentissages domaine aff. et psy-mot.	2	40	1	50	2	1	0	-	6	0
10	Total	-	1	-	7	11	6	11	4 épreuves	19	9 épreuves

Source : Auteur

Nous constatons que 25 questions soit 35,7 % ont été composées conformément aux objectifs pédagogiques et 28 objectifs soit 62,2 % n'ont pas fait objet de l'évaluation. Nous avons remarqué 11 cas répartis dans 6 épreuves de congruence parfaite, 11 cas répartis dans 4 épreuves de congruence satisfaisante et 198 cas répartis dans toutes les 9 épreuves de congruence nulle.

## 5. Discussion

De 70 questions posées dans les 9 cours, 25 soit 35,7 % ont été appariées aux objectifs pédagogiques. En outre, 28 sur 45 objectifs soit 62,2% n'ont pas été mesurés par les questions des examens. Cette insuffisance dans l'évaluation des connaissances des étudiants, constituant un désagrément dans le processus d'enseignement apprentissage, serait due à la négligence ou au manque de conscience professionnelle chez la plupart d'enseignants des universités. Alors qu'il est toujours primordial de savoir quels sont les objectifs de la formation lorsque l'on conçoit une évaluation (ANONYME, 2012, p.24).

Dans la deuxième partie de ce tableau, nous constatons 11 cas de congruence parfaite répartis dans 6 cours, ainsi que 11 cas répartis dans 4 cours pour la congruence satisfaisante. Ces résultats sont essentiellement le fruit des efforts fournis par certains enseignants dans l'établissement d'une adéquation entre les questions d'examen et les objectifs pédagogiques. Pour l'ensemble de ces 9 épreuves semestrielles retenues dans cette étude, la congruence nulle est la plus prépondérante avec 198 cas répartis dans toutes les 9 épreuves, soit 100 %. Au regard de ces résultats, nous pouvons considérer que la majorité de ces épreuves ont été constituées principalement des questions n'ayant pas de relation avec leurs objectifs pédagogiques.

Ce résultat rejoint le propos de GALLUP (1947, p.325) qui affirme que souvent les praticiens (enseignants) ne suivent aucune méthodologie pour construire les questions. Ayant en tête ce qu'ils souhaitent mesurer, ils se fient à leurs intuitions pour produire les questions. Pourtant, il est indispensable d'avoir les idées claires sur plusieurs points avant de se lancer dans cette production. Cependant, bien que MASANDI, A. (2007, p.41) confirme que la pratique d'évaluation des connaissances pose d'énormes difficultés aux enseignants dans leurs activités quotidiennes, il sied de faire remarquer qu'une seule épreuve (Statistique Descriptive) a remplie toutes les conditions pour une validité de l'épreuve avec ; 11 questions soit 92 % appariées aux objectifs pédagogiques, seulement 3 soit 37 % d'objectifs non

mesures par les questions. En outre, 3 cas de congruence parfaite, 5 cas de congruence satisfaisante et seulement 16 cas de congruence nulle.

Cette distinction dans sa manière de concevoir l'évaluation serait expliquée par une bonne conscience professionnelle, connaissance des valeurs morales et citoyennes, un haut sens d'honnêteté et un amour envers son métier.

## Conclusion

Au terme de cette étude qui a porté sur la validité de quelques épreuves du premier semestre à la Faculté de Psychologie et des Sciences de l'Education, il sied de rappeler que notre objectif était de vérifier si les examens du premier semestre au Département des Sciences de l'Education ont été valides. Concrètement, nous avons cherché d'une part, à voir à la fois combien de questions sont appariées aux objectifs pédagogiques et combien d'objectifs pédagogiques ne sont pas mesurés par les questions des épreuves ; d'autre part, puis à dégager le degré de relation le plus prépondérant entre les questions et les objectifs pédagogiques.

Au regard des résultats des études antérieures, nous avons émis l'hypothèse selon laquelle, les épreuves du premier semestre au Département des Sciences de l'Education ne mesureraient pas globalement les objectifs opérationnels formulés lors des enseignements. Autrement dit, la majorité de questions ( $> 50\%$ ) ne s'apparieraient aux objectifs, d'une part, et la plupart d'objectifs ( $> 50\%$ ) ne se mesureraient par les questions, d'autre part. En outre, par rapport au degré de relation Question/Objectif, la congruence nulle serait le degré le plus prépondérant de rapport entre les questions et les objectifs pédagogiques.

Après analyse des données, les résultats obtenus nous ont révélé que la congruence nulle est le degré le plus prépondérant de rapport entre les questions et les objectifs pédagogiques. Explicitement, moins de 50 % de questions ( $35,7\% < 50\%$ ) sont appariées aux objectifs et la plupart d'objectifs ( $62,2\% > 50\%$ ) ne sont pas mesurés par les questions.

Au regard de ces résultats, nous pouvons donc affirmer que notre hypothèse principale, à travers ses explications, a été confirmée. Autrement dit, les épreuves du premier semestre n'ont pas été valides.

Sans vouloir dénoter une quelconque prétention, d'une part, les résultats de cette étude réveillent l'attention des chercheurs en matière de l'évaluation des acquis scolaires, leur montrent les problèmes auxquels est confronté l'évaluation des acquis dans moult systèmes éducatifs et de manière spécifique celui de la République Démocratique du Congo et d'autre

part ils sensibilisent les praticiens (les enseignants) à veiller sur les règles conduisant à bien évaluer.

Fort de ces résultats, quelques suggestions paraissent importantes. Il est donc une urgence de procéder à la formation des enseignants dans la science de construction des questions. Quelques suggestions paraissent importantes et en même temps revient quelques avenues de recherche pour qui voudrait nous emboîter le pas. Au regard de ce qui est précédemment souligné, il est de toute urgence de procéder à la formation des enseignants dans la science de construction des questions. Des recyclages réguliers pourraient en être fructueux. Nous pensons par conséquent que d'autres chercheurs pourront s'inspirer de ces résultats pour aborder des nouvelles questions et pistes de recherches exploitées par nous, soit prolonger la même orientation que la nôtre jusqu'au bout de la démarche que nous avons adoptée.

En dépit des analyses effectuées, cette étude comporte deux limites évidentes. La première tient au fait qu'elle s'est planchée sur une évaluation formalisée de la qualité des questions par des experts, elle susciterait encore plus d'intérêt, si elle était réalisée soit à partir des scores des questions, soit pour dégager le sens à attribuer aux scores aux questions. Des études s'orientant dans ces pistes de recherche serraient les plus attendues. Le fait d'avoir calculé le degré de concordance entre questions et objectifs sans songer à la concordance de la correspondance des appréciations des juges constitue la deuxième limite de cette étude.

## REFERENCES BIBLIOGRAPHIQUES

1. ANDRE, N., et al (2015). *La validité psychométrique : un regard global sur le concept centenaire, sa genèse, ses avatars*. *Mesure et évaluation en éducation*, 37(3), 125–148. <https://doi.org/10.7202/1036330ar>
2. ANGOFF, W. H. (1988). *Validity: an evolving concept*. In H. Wainer & H. I. Braun (Eds.), *Test validity* (p. 19-32). Hillsdale, NJ: Routledge.
3. ANONYME, (2012). *Formation des objectifs pédagogiques, S.l., S.e.* 24p.
4. ANGOFF, W.H., (1988). *Validity: An evolving concept*, Wainer, Braun, 83p.
5. BUCKINGHAM, B. R., et al (1921). *Report of the standardization committee*. *Journal of Educational Research*, 4(1), 78-80.
6. CROCKER, L. & ALGINA, J., (1985). *Introduction aux théories du test moderne et classique*, New York : Holt, 213p.
7. DEPELTEAU, F., (2003). *La démarche d'une recherche en sciences humaines*, Québec, de Boeck, 6p.
8. EBEL, R.L., (1976). *Essentiels de mesure en éducation*, Englewood, Prentice-Hall, 411p.
9. GALLUP, G., (1947). *Mesure en éducation*, New York, McKay, 325p.
10. HAYNES, S.N., et al. (1995). *Content validity in psychological assessment: A functional approach to concept and methods*. *Psychological Assessment*, 7, 238-247. <http://dx.doi.org/10.1037/1040-3590.7.3.238>
11. KANE, M. T. (2013). *Validating the interpretations and uses of test scores*. *Journal of Educational Measurement*, 50(1), 1-73. doi:10.1111/jedm.12000
12. KAMBA, A., (2004). *Validation des épreuves de compréhension de texte et de mathématiques destinées aux élèves de 6<sup>ème</sup> année Primaire de Kisangani*, UNIKIS : *Revue congolaise de Psychologie et de Pédagogie*, Vol.1, pp 41-57.
13. Kelley, T. L. (1927). *Interpretation of educational measurements*. New York, NY: Macmillan.
14. LAUGIER, H et al (1934). *Etude docimologique pour le fonctionnement des examens et concours*, publication du travail humain ; Paris, Maison du livre.
15. LAUGIER, H et al (1936). *Elaboration statistique des données numériques de l'enquête sur correction des épreuves du baccalauréat*, in : *la correction des épreuves écrites dans les examens*, Paris, Maison du livre.

16. MASANDI, A., (2007). *Attitude des élèves de sixième année secondaire des écoles de la ville de Kisangani face à l'évaluation scolaire*, Thèse de doctorat, Inédit, UNIKIS : FPSE, 41p.
17. MESSICK, S. (1989). *Validity*. In r. Linn (Ed.), *Educational Measurement* (p. 13-103). Washington, DC: American Council on Education and Macmillan.
18. NEWTON, P. E. (2012). *Clarifying the consensus definition of validity*. *Measurement: Interdisciplinary Research and Perspectives*, 10(1-2), 1-29. doi:10.1080/15366367.2012.669666
19. NIOZET, G. et al (1961). *Etudes docimologiques sur la correction de l'écrit des baccalauréats*, BINOP.
20. NIOZET, G. et al (1975a). *Identification ou discrimination des énoncés d'un OCM Aix-en-Provence*. Université de Provence, laboratoire de Psychologie expérimentale. Rapport de Recherche.
21. NIOZET, G. et al (1975b). *L'utilisation des critères d'évaluation : Etude par procédure du relais d'information*, Aix-en-Provence. Université de Provence, laboratoire de Psychologie expérimentale. Rapport de Recherche.
22. NIOZET, G. et al (1977). *Note sur l'expérimentation hors laboratoire en docimologie, psychologie française*. Aix-en-Provence. Université de Provence, laboratoire de Psychologie expérimentale. Rapport de Recherche.
23. MUGENOT, L., (2016). *L'évaluation scolaire en question : de l'activité des enseignants aux conduites des élèves*, Inédit, 304p.
24. PIERON, H. (1963). *Examen et docimologie ; tome2*, Paris, Bordas.
25. SCHMIDT, F. L. (2012). *Cognitive tests used in selection can have content validity as well as criterion validity: a broader research review and implications for practice*. *International Journal of Selection and Assessment*, 20(1), 1-13. doi:10.1111/j.1468-2389.2012.00573.x