

# VALORACIÓN DE LOS INSTRUMENTOS DE EVALUACIÓN EN LOS PROCESOS DE APRENDIZAJE

Oscar Flores Pérez<sup>[16]</sup>

Carmenza Chang Herrera<sup>[17]</sup>

Norma Corea Torres<sup>[18]</sup>

## Resumen

El objetivo fue valorar los instrumentos de evaluación de los aprendizajes aplicados a los estudiantes en las carreras regulares de URACCAN, Recinto las Minas, 2008. El procedimiento de valoración fue mediante cinco muestras estratificadas en las carreras: Ingeniería, Administración, Ciencias Sociales e Informática para un nivel de significancia de  $\alpha=0.05$ .

Se realizaron entrevistas y encuestas a estudiantes y docentes así como revisión bibliográfica. Se encontró que se utilizan nueve categorías de ítems en los instrumentos evaluativos y que la planificación es espontánea. Los instrumentos evaluativos más utilizados son las preguntas cortas, respuestas cortas y verdadero – falso.

La concordancia de los instrumentos evaluativos aplicados con las teorías que sostienen el proyecto pedagógico no es exhaustiva. Existe poca relación entre aspectos instruccionales, número de reactivos, representatividad de contenidos y niveles cognoscitivos evaluados y la distribución del grado de dificultad tiende a ser muy fáciles.

El tipo de categoría de ítems más frecuente fue preguntas cortas (61%), la validez de contenido evaluada por estos reactivos fueron influenciados por el nivel de profundidad exigido por la pregunta. No se encontró fórmula alguna para evaluar la confiabilidad para preguntas con ítem de respuestas engendradas, sólo para las de respuestas seleccionadas, pero sobre la base del puntaje asignado por la o el docente a los instrumentos evaluativos analizados las carreras de Ingeniería Agroforestal, Licenciatura en Administración de Empresas y Ciencias Sociales con mención en Desarrollo Local e Ingeniería en Zootecnia se obtuvieron coeficientes alfa de 0.68, 0.48, -0.84 y 0.70, respectivamente.

**Palabras claves:** Instrumentos, evaluación, planificación, aprendizaje, validez y confiabilidad.

[16] MSc. Docencia Universitaria. Docente Tiempo Completo URACCAN, Recinto Las Minas. oscarflores7@hotmail.com

[17] MSc. Docencia universitaria. Docente Tiempo Completo URACCAN, Recinto Las Minas. carmenzachang@yahoo.com

[18] PhD. Tutora. Docente en Maestría Docencia Universitaria. normacandida@hotmail.com

## Introducción

La cosmovisión de cada docente en el proceso evaluativo de la enseñanza–aprendizaje puede imprimir cierto grado de subjetividad y variabilidad en las puntuaciones asignadas a cada estudiante independientemente de la forma de evaluar y en este sentido, la híper o hipo valoración de los tests, encubre muchas veces la inseguridad personal de quien maneja el recurso y conlleva a la empobrecida circunstancia de trastocar lo que es propiamente humano por el resultado de lo técnico.

Previas investigaciones han encontrado que el tipo de producción de respuesta, muestran una alta validez, pero una confiabilidad media, debido principalmente a la subjetividad del evaluador en la corrección. Al respecto una investigación realizada en Estados Unidos identificó cuatro tipos de comentario entre los docentes: alabanza, aceptación, remedio y crítica. Además el favoritismo hacia los varones era mayor en las reacciones más útiles de alabanza, crítica y remedio. Cuando las docentes dedicaban más tiempo y se esforzaban en evaluar específicamente la ejecución de un estudiante, casi siempre se trataba de un varón.

Entonces, la evaluación no tiene sentido por sí misma, sino como resultante del conjunto de relaciones e interacciones entre: objetivos, métodos, modelo pedagógico, estudiante, docente y sociedad. Y, en este sentido tanto la rutina profesional como la innovación apresurada entrañan el grave riesgo de la despersonalización, de ahí, cabe preguntarse, *¿Cuál es la valoración acerca de los instrumentos de evaluación de los aprendizajes en los cursos de la modalidad regular de la URACCAN Siuna, con respecto a las teorías didácticas que sostienen el proyecto pedagógico?*

Con este trabajo se pretende mostrar una panorámica general del proceso de evaluación de los aprendizajes y, de ahí, ofrecer pautas para evaluar no sólo los aprendizajes, sino el proceso de enseñanza, visibilizar conscientemente y con eficacia las fortalezas y los defectos técnicos de las evaluaciones.

## Revisión de Literatura

Desde la URACCAN la evaluación es concebida como un proceso continuo en el cual las diferencias individuales que apunta al desarrollo de habilidades con integralidad, toma en cuenta el desempeño y la reflexión activa y crítica de los estudiantes (URACCAN, 2004, pág. 28).

Sin embargo, para D´Agostino de Cersósimo (2004), evaluar es un hecho de juzgar el valor de algo..., y de atribuirle o negarle grados de mérito y de calidad. Este mismo autor, señala que para emitir juicios evaluativos se requieren, como mínimo cuatro elementos: el objetivo, idoneidad por parte de quien formula el juicio en el campo

específico (conocimiento de lo que se juzga), información relevante y útil para el tipo de juicio que debe formularse además de parámetros de comparación explícitos o implícitos.

Corea (2007) añade que es necesaria también “la determinación de la magnitud cuantitativa o cualitativa de lo que se valora”. Sin embargo, al momento de evaluar, se puede limitar a criterios de grupo (evaluación normativa), a criterios conductuales –evaluación criterial o considerar la personalidad, posibilidades de progreso y limitaciones del propio estudiante.

Se habla de evaluación sumativa para designar la forma, mediante la cual, se mide y juzga el aprendizaje con el fin de certificarlo, asignar calificaciones, determinar promociones, etc. y, cuyos objetivos según D´ Agostino de Cersósimo (2004) son calificar en función de un rendimiento, otorgar una certificación según los indicadores establecidos como mínimos para el curso o etapa.

## **Planificación de las evaluaciones**

En este sentido, las pruebas elaboradas por el personal docente, según De Flores (2000), deben poseer algunas características fundamentales como.... objetivos de la prueba, instrucciones claras para desarrollar la prueba, ítems claros, precisos y concisos, medir el aprendizaje de un aspecto con varios tipos de ítems, abarcar los objetivos que se pretenden evaluar, formatos adecuados en la elaboración de los ítems y procurar pruebas con niveles de validez y confiabilidad apropiados.

La evaluación implica juicios de valor que deben sustentarse en juicio de hecho y de verdad, (López, 2007) la valoración es un proceso ontológico de elementos analíticos, sintéticos, objetivos y subjetivos que pueden participar o no en la evaluación.

## **Tipos de instrumentos de evaluación**

En las pruebas de elaboración o respuesta construida se le pide al estudiante que tras recordar una información determinada, la elabore, la organice y la maneje de una manera concreta. Entre estos tipos de instrumentos de evaluación se hallan: los ensayos o preguntas largas, las preguntas cortas, la entrevista de evaluación, la evaluación de portafolios, los trabajos de curso, el mapa conceptual y las pruebas de ejecución.

## **Concordancia de los instrumentos con las teorías didácticas que sostienen el proyecto pedagógico**

Los principales componentes curriculares de una planificación didáctica en el paradigma cuantitativo, corresponden a: objetivos, contenidos, estrategias de aprendizajes y evaluación; De Flores (2000) considera que entre estos componentes debe haber una

estrecha correlación que propicien el desarrollo de las capacidades propuestas en los objetivos generales del perfil profesional.

Los instrumentos de medición, utilizados para la evaluación del rendimiento escolar, deben ser equilibrados y reflejar objetivos evaluables. Además, deben cumplirse con las normas de diseño, elaboración y aplicación o ejecución de cada instrumento evaluativo –de selección o de construcción–, con conocimientos de sus ventajas y desventajas.

### **Grados de dificultad de los ítems de los instrumentos de evaluación aplicados**

El Índice de dificultad ( $p_i$ ) De Flores (2000), para las pruebas referidas a normas, los mejores reactivos son los que tienen índices de dificultad de alrededor de 0.50

Es una proporción que va desde 0 a 1, con una escala que va de muy fácil (81 a 100), fácil (61 a 80), moderadamente fácil (41 a 60), difícil (21 a 40) hasta muy difícil (00 a 20). (Abad, et al., s.f). Lo ideal es que los ítems no sean ni demasiado fáciles ni demasiado difíciles.

### **Validez y confiabilidad de los instrumentos de evaluación**

Según Muñiz, Roger y Swaminathan (1989), las propiedades de los ítems dependen del tipo de competencias o dominio de los sujetos sobre la variable medida y del conjunto de ítem (pág. 83).

Aquí, Woolfolk, (1999) supone que los buenos reactivos son los que responden bien los estudiantes que tuvieron un buen desempeño en todo el examen y, viceversa y, este desempeño puede ser interpretado atendiendo a la siguiente escala: Muy Alto (0,81 a 1,00), Alto (0,61 a 0,80), Moderado (0,41 a 0,60), Bajo (0,21 a 0,40) y Muy Bajo (0,0 a 0,20) (Backhoff, Larrazolo, & Rosas, 2000, pág. 17)

Si  $D_i = -0.20$  a  $-1.00$  ello significa que existe una fuerte discriminación negativa entre los estudiantes con calificaciones altas y bajas ¿está mal la clave?, ¿permite dos respuestas correctas?, etc.). (Woolfolk, 1999, pág. 563).

### **Comportamiento de distractores**

En todo reactivo de selección múltiple cada uno de los distractores debe diferir de la opción correcta en un solo aspecto y, Solano (2003), un buen distractor es aquel que se aproxima inexactamente a la opción correcta, para ello según Medina & Verdejo (2000) se espera que como mínimo 10% de las personas los seleccionen y que sea seleccionado con mayor frecuencia por los estudiantes con puntuaciones bajas.

Para Hernández & Fernández, (2006) la validez de contenido se refiere al grado en que un instrumento refleja un dominio específico de contenido –representatividad– de lo que se mide. En este sentido, hay dos tipos de validez de contenido: la validez superficial que se refiere al grado en que se está midiendo según la valoración subjetiva e intersubjetiva de los educadores y del instrumento de medición (Ruiz, 2008) y, la validez muestral que se refiere a que el instrumento contenga una muestra representativa del universo de la materia de interés (Namaskforoosh, 2007, pág. 67).

## Materiales y Métodos

La investigación se realizó en la Universidad de las Regiones Autónomas de la Costa Caribe Nicaragüense (URACCAN) ubicada entre las coordenadas geográficas 13°44' de latitud norte y 84°46' de longitud Oeste en la RAAN, Nicaragua, específicamente en las carreras de Ingeniería en Zootecnia y Agroforestal y, las Licenciaturas en Administración de Empresas, Informática Administrativa y Desarrollo Local, todas en la modalidad regular, Siuna 2008.

El tipo de estudio es descriptivo transversal con enfoque cuantitativo. El universo lo constituyeron las formas de evaluación de aprendizajes y la muestra se calculó en función del número de asignaturas y sobre la base de la fórmula de estimación de  $n$  para variable cuantitativa finita y para un muestreo estratificado.

$$n = \frac{\sum_{i=1}^k N_i p_i q_i}{ND + \frac{1}{N} \sum_{i=1}^k N_i p_i q_i}$$

Donde,  $N_i$  = número de pruebas realizadas en la carrera,  $n$  = muestra total y  $D = B^2/4$  y  $B$  es el límite del error de muestreo.

$N = 90$ , nivel de confiabilidad = 95%,  $p = q = 0.5$  y luego,  $n = 73$  asignaturas.

Las fuentes primarias fueron las pruebas escritas, docentes y estudiantes y, las secundarias, cuaderno de registro del docente, libros básicos y complementarios.

Las técnicas e instrumentos de recopilación de información fueron entrevistas, encuestas, grupo focal y análisis de contenido.

$$\alpha = \frac{K}{K-1} \left[ 1 - \frac{\sum S_i^2}{S_r^2} \right]$$

La confiabilidad de los instrumentos evaluativos se analizó a través del coeficiente Alfa de Crombach, cuya fórmula es:

La validez de contenido se determinó mediante un análisis cualitativo exhaustivo de aquellos elementos vinculados con las pruebas evaluativas aplicadas: representatividad de los contenidos impartidos (validez muestral) y la congruencia entre prueba y objetivos, el aspecto exterior de la prueba, el atractivo y el interés que despierta en aquellos a los que va a ser aplicada –validez didáctica– y especialmente lo que se refiere a qué tan exitosamente un instrumento representa y mide un concepto teórico.

Para analizar el grado de dificultad, discriminación y confiabilidad se tomó un examen al azar por carrera.

$$\text{Fórmula: } p_i = A_i/N_i$$

Donde;

$p_i$  = Índice de dificultad del reactivo i.

$A_i$  = Número de aciertos en el reactivo i.

$N_i$  = Número de aciertos + número de errores en el reactivo i.

En el análisis de datos se siguió la estrategia de triangulación, de modo que se solapen técnicas de análisis de contenido y de entrevistas con preguntas estructuradas y en profundidad junto a estrategias de investigación evaluativa.

### **Cálculo matemático más simple sugerido por (Medina & Verdejo, 2000, p . 249):**

Ordene descendientemente las puntuaciones. Separe en dos grupos: altas y bajas puntuaciones. Cuando el grupo es pequeño puede trabajarse con la mitad del grupo (1/2), y 2/3 cuando el grupo es numeroso.

$$\text{Fórmula: } D_i = 2 * (NAGR - NAGI) / N$$

Donde:

$D_i$  = índice de discriminación del reactivo i

NAGR = Número de aciertos en el reactivo i de las personas con las puntuaciones más altas en el test

NAGI = Número de aciertos en el reactivo *i* de las personas con las puntuaciones más bajas en el test

N = número total de estudiantes que contestaron el ítem en ambos grupos.

**RESULTADOS Y DISCUSIÓN**

**Planificación de instrumentos evaluativos**

El proceso de planificación evaluativa en la universidad URACCAN Siuna, es espontánea y consiste fundamentalmente en valorar los objetivos del programa de asignatura, los objetivos de unidad del programa y a partir de ello se elaboran los ítems o preguntas que luego se constituyen en los instrumentos evaluativos a aplicarse y, con respecto a la distribución de puntajes, esto generalmente se hace con referencia las orientaciones evaluativas de cada programa de asignatura. Este proceso es el resultado de compartir experiencias con otros docentes, las características de la asignatura que cada quien facilita y la normativa vigente en la universidad URACCAN.

En términos generales, los docentes expresaron haber recibido capacitaciones sobre el diseño de las pruebas, pero no sobre planificación evaluativa, excepto un docente, quien indicó conocer dos maneras de realizar una planificación evaluativa: las tablas de especificaciones y el proyecto de prueba.

En la planificación evaluativa se establece la concordancia y secuencia lógica entre los objetivos, la metodología aplicada en el proceso de enseñanza y los instrumentos evaluativos. Además los estudiantes coincidieron en que la estructura de las pruebas está acorde con el perfil de la carrera y a la temática; pero, a pesar de existir derivación de objetivos entre los instrumentos de evaluación y programas de asignaturas, existe dificultad para la redacción correcta de los mismos, esto último no concuerda con Alvarado (1995).

**Cuadro No. 1.** Dificultades encontradas en el proceso evaluativo.

Carrera	No se encontró	
	Syllabus	Proyectos de examen
Agroforestal	5	0
Administración de empresas	0	3
Desarrollo local	2	3
Zootecnia	3	2
Informática	2	1

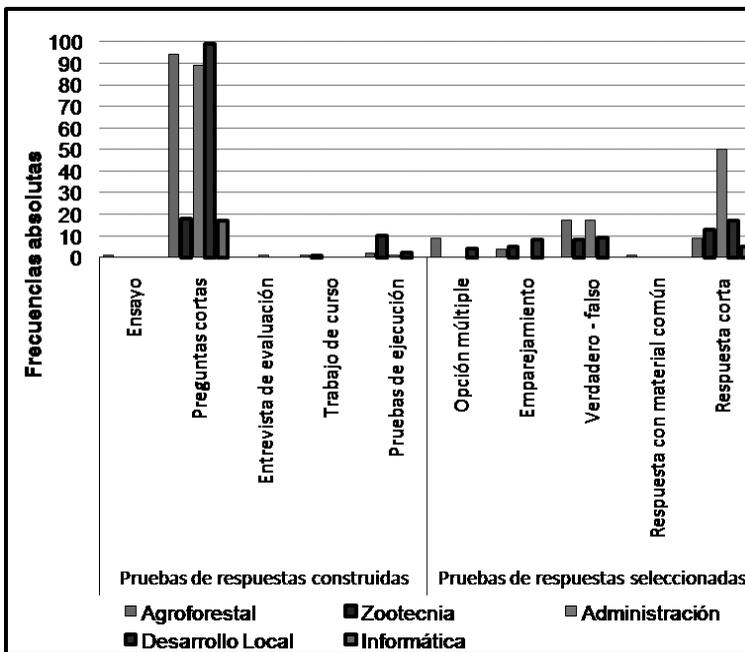
Se constató que existe un proceso de planificación que concuerda con López (2007) en el cual se evidencia la secuencia derivada de los objetivos de programa y los descritos en los instrumentos evaluativos, esto es bueno, lo mismo ocurre con los contenidos. Sin embargo, no constituyen una muestra significativa del programa y contenidos desarrollados, además tienen tendencia a sólo evaluar el nivel cognitivo, obviando así, otros muy importantes: los procedimentales y actitudinales.

Aunque no se refirieron al orden de presentación de preguntas en el instrumento, en las pruebas se notó dificultad en el orden recomendado por Alvarado (1995) orden creciente de complejidad. Además, los instrumentos evaluativos muestran diferentes niveles de estructuración, que resulta poca precisa su practicabilidad, objetividad, confiabilidad y validez, esto es debido a que, el personal docente, ha ido adoptando y modificando modelos que han sido compartidos por otros docentes, sin conocer a profundidad, las ventajas y desventajas de los mismos.

### Tipos de instrumentos de evaluación sumativa

De la revisión de documentos (73 proyectos de instrumentos evaluativos) de curso regular en la universidad URACCAN Siuna, se obtuvieron los siguientes hallazgos. ver gráfico No.1.

Gráfico No. 1. Instrumentos evaluativos por carrera utilizados en la Universidad URACCAN, Siuna.



En la carrera de Ingeniería Agroforestal se identificaron nueve categorías de ítems distribuidos principalmente en las preguntas cortas con 68 por ciento de aparición entre los instrumentos evaluativos y los ítem de verdadero – falso, con 12.3 por ciento. En Ingeniería en Zootecnia se presentaron los instrumentos evaluativos con las frecuencias más equilibradas: las preguntas cortas (33%); las respuestas cortas, (24%); las pruebas de ejecución (18%), entre las otras opciones. Cabe señalar que los trabajos de curso son los menos frecuentes y son considerados difíciles, pero excelentes, dado que la teoría se valida en el campo y se retroalimenta con más información.

Los trabajos de curso en la Universidad URACCAN, Siuna, son evaluados por el o la docente de la asignatura, pudiendo el titular de la asignatura invitar a otras personas para validar los informes. Para otorgar la calificación se toman en cuenta los siguientes aspectos:

- La calidad del trabajo (cumplimiento de los objetivos planteados)
- La creatividad o independencia del estudiante.
- La originalidad.
- La estética del trabajo.
- La calidad de la presentación y defensa, en caso exista.
- Pertinencia.

En Administración de Empresas se verificaron cinco categorías de ítems, preguntas cortas con un 56.3 por ciento; las respuestas cortas, con 31.6 por ciento y, verdadero – falso con 10.7 por ciento; en Desarrollo Local, seis categorías de ítems, preguntas cortas (71%) seguido de respuesta corta con (12%), y en Informática Administrativa se evidenciaron dos categorías de ítems, las preguntas cortas con (77%) y las respuestas cortas con (23%).

Independientemente de las formas evaluativas que se utilicen para evaluar aprendizajes educativos, es importante destacar el merecido cuidado que se ha de tener en la observación, la medición, la cuantificación y el control en todo el proceso, especialmente en el diseño del instrumento evaluativo y, lo importante es que la preparación de las pruebas se fundamente en los objetivos específicos de la asignatura, de tal manera que exista total concordancia entre los objetivos desarrollados y la prueba a aplicar, además, el programa de evaluación debe incorporar aquel tipo de prueba que mejor logre expresar los resultados del aprendizaje.

### **Concordancia de los instrumentos evaluativos con las teorías didácticas**

En la carrera de Ciencias Sociales con Mención en Desarrollo Local, existe correspondencia entre los contenidos evaluados y los contenidos del programa de asignaturas, Sin embargo, el 79 por ciento de los proyectos de examen de las asignaturas revisadas

en la carrera, sólo evaluaron contenidos conceptuales y únicamente el 21 por ciento hace una combinación de contenidos procedimentales. Aunque el 71 por ciento de los reactivos corresponden a la categoría de preguntas cortas, su uso no fue de máximo provecho, solo se refieren a definición de conceptos. Los conceptos son importantes, pero no constituyen lo más valioso del aprendizaje, se requieren habilidades manuales y raciocinio, coherentes entre objetivos, contenidos, estrategias de aprendizajes y evaluación.

**Cuadro No. 2.** Concordancia de los instrumentos con las teorías didácticas

<b>Carrera</b>	<b>Categorías de ítem</b>	<b>Claridad en redacción</b>	<b>Instrucciones generales</b>	<b>Instrucciones específicas</b>	<b>Esencial</b>	<b>Contenido desarrollado</b>	<b>Contenido evaluado</b>	<b>Omisión de contenido</b>	<b>Objetivos desarrollados</b>	<b>Objetivos evaluados</b>	<b>Omisión de objetivos</b>
Agroforestal	1.96	.89	.00	.87	.93	19.44	7.87	59.53	12.24	4.81	60.70
Administración	1.89	.63	.00	.89	.79	19.53	7.11	63.61	9.00	3.32	63.16
Desarrollo Local	1.39	.67	.00	.61	.75	15.44	6.17	60.05	5.00	2.83	43.33
Zootecnia	2.00	.94	.00	.81	.88	22.20	8.38	62.27	12.50	5.67	54.67
Informática	1.00	.50	.00	.50	.67	20.00	6.33	68.33	13.33	4.33	67.50

La relación entre objetivo de programa y de los instrumentos evaluativos es buena, sin embargo, se excluyen objetivos y contenidos relevantes del programa, lo cual reduce la validez del instrumento. Además, casi siempre se abusa de los mismos tipos de ítems obviando así la utilidad de los otros, con lo cual se evaluarían otros niveles de conocimiento que coadyuvarían de mejor manera con la preparación profesional o especialidad.

En términos generales, se evidenció poca pericia en la preparación y planeación de cada una de las formas de instrumentos evaluativos aplicados, entre las cuales podemos señalar instrucciones imprecisas, preguntas con niveles de dificultad inapropiados, pruebas demasiado cortas (no representativas) e ítems incongruentes con el dominio o universo de contenido, esto es, instrumentos de medición pocos equilibrados y pocos exhaustivos. Esto es, no abarcan satisfactoriamente una muestra representativa de los conocimientos, las habilidades y las actitudes producto del aprendizaje que se evaluaron durante su aplicación.

En Ingeniería Agroforestal, el nivel de dificultad promedio del instrumento fue de 0.58 lo cual se clasifica como un instrumento de mediana dificultad, el índice de discriminación (ver gráfico 2 en la página siguiente), de 0.33 para una clasificación de bueno con posibilidades de mejorarlo.

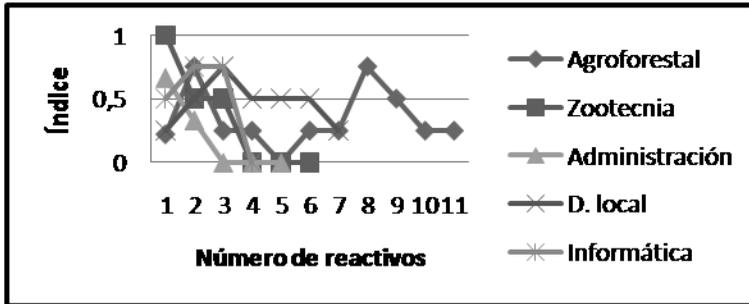


Gráfico No. 2. Índice de discriminación por carrera.

En términos generales, el instrumento evaluativo analizado, para Ingeniería Agroforestal y la licenciatura en Ciencias Sociales, cumplen con el nivel medio de dificultad (0.58), sugerido,  $\pi = 0.5$  y; con respecto al índice de discriminación las carreras de Ingenieras Agroforestal y Zootecnia y la licenciatura en Ciencias Sociales los instrumentos presentaron valores aceptables de discriminación-0.31, 0.33 y 0.46, respectivamente.

En el caso de la Licenciatura en Administración de Empresas se mostró un valor de discriminación igual a 0.20, aunque es un valor que está dentro del rango aceptable, merece revisión de cada uno de los reactivos, no solo el aspecto técnico del instrumento evaluativo – propiedades de los ítems, número de ítems -, sino también las características del grupo, las habilidades de su trasfondo educativo y el tipo de instrucción que recibieron los estudiantes, entre otros aspectos a considerar.

### Validez y confiabilidad de los instrumentos evaluativos

En gráfico No. 3 se tienen los resultados del comportamiento de los contenidos desarrollados frente a los evaluados.

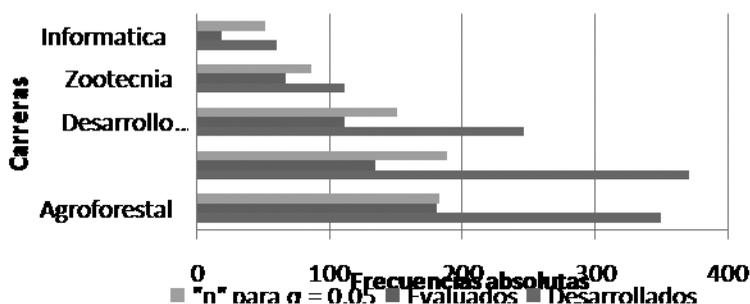


Gráfico No.3. Balance de contenidos por carrera.

Las carreras de Ingeniería Agroforestal, Administración de empresas, Ciencias Sociales, Ingeniería en Zootecnia y la licenciatura en Informática Administrativa, los ítems revisados tienen una relevancia del 93, 79, 75, 88 y 67 por ciento, respectivamente, lo cual es positivo para la validez de contenido. En otro sentido, los contenidos evaluados aparecen en el programa de asignatura, y fueron impartidos en clase; sin embargo, esta validez es afectada por la no representatividad de contenidos evaluados, debido a que las carreras de Administración de empresas, Desarrollo local, Zootecnia e Informática Administrativa omitieron el 63.6, 55.1, 39.6 y 68.3 por ciento, respectivamente.

Para un nivel de significancia  $\alpha = 0.05$ , asumiendo normalidad en el número de contenidos desarrollados y para un muestreo al azar simple se obtuvieron los valores de la gráfica 7. De ella se deduce que la carrera de ingeniería agroforestal mostró buena representatividad en términos de número de contenidos.

En términos generales, se encontraron problemas de claridad – ambigüedad- en la redacción de los reactivos, pero en menor cuantía, no así, en la presentación del instrumento evaluativo, no se describieron instrucciones generales, que indicaran lo que interesaba medir – los objetivos -, con especificaciones acerca de los criterios de valoración.

Otra de las dificultades, el instrumento evaluativo contiene poco ítem, algunos sólo tienen dos categorías de ítem con pocos reactivos, esto limita la inclusión de más contenidos importantes que podrían ser objetivo esencial de evaluación, a esto se añade que los reactivos propuestos independientes de la categoría a la que pertenezcan, por lo general, no permiten la inclusión de contenidos no explicitados en la prueba.

Adicional a lo anteriormente expresado, la calificación es afectada por la subjetividad del docente –tendencia a asignar puntaje arbitrariamente para no reprobar a los estudiantes -. Sin embargo, los juicios de valor deben sustentarse en juicio de hecho

y de verdad, y se debe evaluar en función del aprendizaje que se está midiendo. Un estudiante expresó: “a veces siento como que algunos docentes no se toman el tiempo suficiente para revisar el examen. Cuando reclamamos, acerca de la calificación, sentimos que no somos escuchados. Nosotros comprendemos a los docentes, pero también necesitamos ser comprendidos”

Con referencia a la confiabilidad de los instrumentos evaluativos analizados para las carreras de Ingeniería Agroforestal, Administración de empresas y Ciencias Sociales e Ingeniería en Zootecnia tenemos: 0.68, 0.48, -0.84 y 0.70, respectivamente. En el caso del valor inferior a cero, se puede afirmar que el material de instrucción evaluado es heterogéneo, pero sus ítems no se correlacionan positivamente entre sí, aunque el valor estimado es alto, es negativo, por tanto, debe revisarse el marco muestral de los contenidos evaluados, readecuar los reactivos porque puede existir redundancia entre los reactivos, y se debe aumentar el número de ítem para reducir la heterogeneidad de las variables entre sí.

En las áreas administrativas, los estudiantes manifestaron sentir que la puntuación que reciben es perjudicada. Al respecto, uno de los estudiantes expresó: “uno se esfuerza por salir bien en clase y se gana los puntos asistiendo a clase. Hay docentes que dan diez puntos por asistir regularmente a clase, pero al momento del examen hay estudiantes que reprueban y les dan los diez puntos aunque no hayan asistido regularmente y no hayan participado activamente en clase, sólo para aprobarlos con el mínimo” o bien, lo siguiente, “¡ah sos necio!, entonces, aunque uno participe activamente en clase, esta participación no es tomada en cuenta en la evaluación. A veces te ponen una nota..... te están robando” De acuerdo con López (2007) la evaluación implica juicios de valor, pero los juicios de valor deben sustentarse en juicio de hecho y de verdad, y Márquez (s.f.) se debe evaluar en función del aprendizaje que se está midiendo.

## Conclusiones

El proceso de planificación de instrumentos evaluativos en la universidad URACCAN Siuna, es espontánea y, por lo general, obedece a la jerarquía de objetivos curriculares de los cuales son transformados posteriormente en los ítems que constituyen los instrumentos de evaluación, en cuyo proceso existe dificultad en la redacción de los objetivos de los instrumentos evaluados y las instrucciones de los mismos.

Los instrumentos evaluativos más utilizados son: pruebas cortas, las de respuestas cortas y las de verdadero – falso. En las carreras de Ingeniería en Zootecnia y Agroforestal se presentaron el mayor número de categorías de ítems (9 categorías).

Existe buena concordancia entre los instrumentos evaluativos, los objetivos de programa y syllabus, los contenidos, el número de categorías que debe contener un instrumento evaluativo, pero no es tan buena en aspectos instruccionales, número de reactivos, representatividad de los contenidos y niveles cognoscitivos evaluados, igualmente el grado de dificultad de los instrumentos evaluativos, en promedio oscila entre medianamente fácil a muy fáciles, y el índice de discriminación tiende a ser bajo.

La representatividad de contenidos evaluados se debe mejorar al igual que el grado de dificultad y discriminación. La confiabilidad tiende a ser alta en aquellos instrumentos, que presentan mayor número de reactivos y mejor distribución del grado de dificultad, pero los resultados son afectados por la subjetividad y el enfoque dado a cada uno de los reactivos.

## Lista de Referencias

Corea, N. (2007). Evaluación Educativa. *Documento Básico de Evaluación Educativa 110*). Managua: Universidad Nacional Autónoma de Nicaragua.

D'Agostino de Cersósimo, G. (2004). *Aspectos teóricos de la evaluación educacional. Una orientación para su puesta en práctica en la enseñanza primaria*. (1ª ed.). San José, Costa Rica: EUNED.

De Flores, M. (10 de 07 de 2000). *Postgrado en Docencia Superior*. (U.Panamá)

Recuperado el 1 de Agosto de 2008, de Sistema de evaluación: <http://www.geocities.com/eds607/DOC3.htm>

Hernández, R., Fernández, C., & Baptista, P. (2006). *Metodología de la investigación* (4ª ed.). Iztapalapa: McGraw-Hill.

López, M. (2007). *Planeación y evaluación del proceso enseñanza aprendizaje: Manual de docente*. (1ª ed.). México: Trillas.

Medina, M., & Verdejo, A. (2000). *Evaluación del aprendizaje estudiantil*. Recuperado el 05 de 07 de 2009, de <http://www.agapea.com/libros/Evaluación-isbn-842770464x-i.htm>

Muñiz, J., Rogers, J., & Swaminathan, H. (1989). Robustez de las estimaciones de modelo de Rash en presencia de aciertos al azar y discriminación variable de los ítems. *Anuario de Psicología*, 82-97.

Solano, G. (2003). *Diseño lógico de exámenes*. México: Trillas.

URACCAN. (3 de septiembre de 2004). *Modelo Pedagógico de URACCAN*. Managua, Nicaragua. *estudiantil y de la culminación de los estudios*. Managua, Nicaragua.

Woolfolk, A. (1999). *Psicología educativa* (7ª ed.). México: PRENTICE HALL.