

NUEVO METODO PARA LOGRAR LA OPTIMA CALIFICACION DEL TEST DE ENSAYO

Sergio Domingo Ortíz Martínez *

Sumario

Se encontró un método que permite lograr una alta objetividad, facilidad y rapidez, evaluación del desarrollo de un problema, asignación de una ponderación real, y una alta precisión en los cálculos durante la calificación del test de ensayo restringido. El trabajo se basó en el uso de una calculadora programable como ayuda en la realización de los cálculos.

La variante al sistema de calificación normal es el de establecer, junto con la clave de respuestas esperadas, el número de aciertos correspondiente a cada pregunta (con esto se transforma la respuesta escrita en números sencillos) que luego se podrán procesar en una calculadora programable. La nota que obtiene un estudiante no es más que la sumatoria de los productos de aciertos por el factor de calificación para cada pregunta teniendo las siguientes columnas: número de pregunta, porcentaje de respuestas buenas, regulares, malas, y su varianza. Luego de esto aparece el valor del índice de confiabilidad, error típico de medida, balance del examen y a continuación una tabla estadística que contiene: intervalo de clase, punto medio, frecuencia, frecuencia acumulada y luego los valores de la media, mediana y moda, así como cuartil 1, cuartil 3, número de casos y a continuación una tabla que establece la curva de frecuencias ya normalizadas y su prueba de chi cuadrada de semejanza de frecuencias.

La información anterior se obtiene en un tiempo de 5 minutos luego de haber introducido en la calculadora el número de aciertos conseguidos por los estudiantes.

Introducción

El trabajo docente enfrenta continuamente la búsqueda de técnicas que permitan hacer de la evaluación una

fuerza eficaz de información en el proceso enseñanza-aprendizaje.

Las ventajas (la medida de productos complejos del aprendizaje, la integración y aplicación de las habilidades del pensamiento) (1) que presenta el test de ensayo restringido o semiestructurado como una herramienta para la evaluación, se ve limitada debido a que sus desventajas son de gran importancia, por ejemplo, el que no exista confiabilidad en la calificación, ya que resulta ser una calificación muy subjetiva y de muy difícil ejecución (1). El tiempo para calificar el test de ensayo, es otro de los factores por el cual su uso se ve limitado.

La inclinación para la realización de pruebas de evaluación se ha dirigido hacia el test de selección múltiple, ya que éste presenta la gran ventaja de contar con una evaluación objetiva (sólo se espera una respuesta buena o mala) (1,2) y el sistema de calificación se puede simplificar hasta el grado de que una lectora de tarjetas pueda efectuarla. Las desventajas que presenta el test de selección múltiple: difícil elaboración, sólo exige una acción de reconocimiento o recuerdo, y que puede llegar a influenciar los métodos de estudio de los alumnos, de manera que éstos centren su aprendizaje en la memorización de hechos y descuiden una comprensión más alta y profunda (2), hacen que el uso de este sistema de evaluación provoque un dilema entre una fácil calificación y una respuesta por memorización.

En los cursos donde el proceso de enseñanza-aprendizaje requiere que los alumnos manifiesten creatividad, no se puede aplicar el test de selección múltiple y el uso del test de ensayo restringido se hace necesario. Debido a esto, y a todas las ventajas que contrae el test de ensayo semiestructurado, se hizo necesario encontrar un método por el cual se pueda optimizar la calificación de éste, esperando entonces una fácil calificación, objetividad, y un sistema por el cual se pueda aplicar las diferentes informaciones ya existentes (índice de confiabilidad, grado de dificultad, y tipo de medida) que puede proporcionar un examen, invitando el menor tiempo posible.

La información obtenida puede hacer de la evaluación un proceso retroactivo permitiendo evaluar tanto el conocimiento general adquirido por los estudiantes como un índice para el profesor de lo que no se llegó a comprender

* Lic. en Química. Escuela de Química, Facultad de Ciencias Químicas y Farmacia.

por el grupo en general, sobre el motivo de la evaluación. Como se podrá ver, el examen aporta datos de los alumnos y del profesor.

Materiales y métodos

Materiales:

El estudio fue realizado durante los cursos de Química Orgánica II para las carreras de Químico Farmacéutico, Químico Biólogo y Biología durante la Escuela de Vacaciones del mes de Junio de los años 1980, 1981, 1982 y 1983 y cursos de Química Orgánica III para las carreras antes mencionadas en los años 1981 y 1983.

a) Diseño de exámenes

De los 40 exámenes realizados, se diseñaron 23 formatos diferentes, siguiendo las recomendaciones para la redacción de cada pregunta dadas por Lemus L.A. (2):

1. Que el test sea estructurado (Uso de la tabla de especificaciones), con lo cual se asegura la validez del examen.
2. Indicar claramente al estudiante dónde y qué debe hacer.
3. Que las preguntas sean de respuesta limitada.

b) Calificación de examen

Se procedió a calificar de acuerdo a las recomendaciones dadas por Goring (1):

1. Al calificar el examen, se debe establecer la clave de calificación (esbozo claro y completo de las respuestas esperadas).
2. Se debe calificar serie por serie, pregunta por pregunta para así asegurar una máxima objetividad.

c) Cálculos a efectuar

1. Nota de examen:

Se diseñaron y comprobaron programas en lenguajes Asembler y Basic para calculadoras programables marca Casio FX 502P y FX 702P respectivamente. El programa escogido se llamó "CALEXOP 3" (ver Fig. 1, programa 1)

Además de la calculadora programable Casio FX 702P, se utilizó una miniimpresora marca Casio modelo F10 y un adaptador para grabar en cassette marca Casio, modelo F1.

2. Grado de dificultad:

El cálculo se realizó dentro del programa "CALEXOP 3" subrutina indcon 3 líneas 6 (ver Fig. 2, programa 2), según la fórmula propuesta por Downie para test potentes (3):

grado de dificultad (p) = No. Rc / No. ex.

Donde

No. Rc = No. de los que respondieron correctamente.
No. ex = No. de los que realizaron el test.

3. Varianza de cada pregunta:

Se realiza dentro del programa "CALEXOP 3" "línea 7 (ver Fig. 2, programa 2)" por la fórmula siguiente, propuesta por Downie (3):

$$\text{Varianza} = p \times q$$

Donde

p = grado de dificultad

q = 1 - p

FIGURA No. 1

```

P0: 275 STEPS
  5 PRT "PARCIAL Q
    FINAL":;INP E$
 10 PRT "# ";INP A
    $
 15 G$="CURSO"
 20 PRT G$;:INP B$
 25 H$="CARRERA"
 30 PRT H$;:INP C$
 35 K$="FECHA"
 40 PRT K$;:INP D$
 45 WAIT 20
 50 MODE 7: IF E$="F
    INAL" THEN 65
 60 PRT " ";A$;"
    ";E$;GOTO 67
 65 PRT " EXAMEN
    ";E$
 67 PRT
 70 PRT G$;" ";B$
 75 PRT
 80 PRT H$;" ";C$
 85 PRT
 90 PRT K$;" ";D$
 95 PRT :PRT
100 PRT "# NOT
    A B "
120 MODE 8:LOAD #0

P2: 342 STEPS
  4 VAC :SAC :INP "
    N",N
  5 FOR I=1 TO N:PR
    T "F";I;"B,R";
  6 INP A(I)
  7 NEXT I
  8 B=0:R=0:FOR I=1
    TO N:X=I:GSB 2
    9
  9 PRT "PRG ";I;:I
    NP K
 10 IF K<0 THEN 20
 11 IF K>S THEN 14
 12 IF K=0 THEN 15
 13 M=I:GSB 28:GOTO
    17
 14 M=I*E6:GSB 28:R=
    R+I:GOTO 16
 15 M=I*E3:GSB 28:R=
    R+0.5
 16 B=B+K*Q/100
 17 NEXT I:MODE 7
 18 PRT CNT+1;"");:C
    SR 3:RAND(B,-2);
    CSR 12:R+J;:MOD
    E 8
 19 STAT R:GOTO 8
 20 LOAD #0
 28 Y=N+I:A(V)=A(V)
    +M:RET
 29 L=A(X)/I*E3:O=FR
    AC L*I*E3:L=INT
    L/I*E3:S=FRAC L*
    I*E3
 30 Q=INT L:RET
 31 SAC :FOR I=1 TO
    N:A(I+N)=0:NEX
    T I:GOTO 8
    
```

4. Indice de confiabilidad:

El valor se consigue aplicando la fórmula No. 20 de Kuder-Richardson (3,4) en el programa "CALEXOP 3" subrutina indcond 3" línea 16 (ver Fig. 2, programa 2):

$$r_{vv} = k/k-1 \times 1 - (\text{sumatoria } p \times q) / S^2$$

Donde

r_{vv} = Coeficiente de confianza
 k = número de preguntas
 S^2 = varianza del test
 $p \times q$ = varianza de las preguntas

5. Error típico de medida:

Se realizó el cálculo dentro del programa "CALEXOP 3" subrutina indcon 3, línea 19 (ver Fig. 2, programa 2) de acuerdo a la fórmula, propuesta por Downie: (3)

$$Se = S \times \text{raíz cuadrada } (1 - r_{vv})$$

Donde

Se = error típico de medida
 S = desviación típica del test
 r_{vv} = coeficiente de confiabilidad.

6. Balance del examen:

El balance del examen, se calculó haciendo la sumatoria del número de las varianzas mayores de 0.55 menos las menores de 0.55 de cada pregunta y multiplicada por 100. Esto se hace en el programa "CALEXOP 3" línea 23 (ver Fig. 2, programa 2).

7. Análisis estadístico de resultados:

Se realiza en el programa "CALEXOP 3" subrutina "TABSTAT" de acuerdo a las fórmulas corrientes para calcular la media, mediana, moda, desviación estándar, cuartiles, en valores agrupados, según Antonio Cerezo Ruiz (5).

8. Normalización de las frecuencias:

Se utiliza el procedimiento estadístico empleado por Antonio Cerezo Ruiz (5). Subrutina "Tabstat"

9. Prueba de la bondad de ajuste para las frecuencias normalizadas.

Se realiza en el programa "CALEXOP 3" subrutina "chicuada", según el procedimiento estadístico indicado por Antonio Cerezo Ruiz (5).

Método:

El método encontrado es el siguiente:

1. Establecer la clave de calificación.
2. Asignar el número de aciertos a cada pregunta (números sencillos).
3. Ponderación de cada pregunta en base de su importancia.
4. Calificar serie por serie.
5. Uso de calculadora programable (programa en basic "CALEXOP 3" (Fig. 1, programa 1) para operar los aciertos conseguidos por los estudiantes en cada examen y así obtener su nota, y luego la información estadística del grupo.

FIGURA No. 2

```

P0: 456 STEPS
3 Y=0:P=0:K=0:MOD
E 7
4 PRT CSR 4:"B
R M VAR"
5 FOR I=1 TO N:X=
N+I:GGB 29
6 G=100/CNT:V=RND
(Q*G,-1):W=RND(
S*G,-1):T=RND(O
*G,-1)
7 Z=(Q+S/2)/CNT:U
=1-Z:A(I)=RND(Z
*U,-4):Y=A(I)+Y
8 WAIT 35:PRT I:"
):CSR 3:V:CSR
5:W:CSR 8:T:CSR
12:A(I)
9 IF Z>0.55 THEN
11
10 K=K+1:GOTO 12
11 P=P+1
12 NEXT I
13 IF CNT>15 THEN
15
14 E=SDX:GOTO 16
15 E=SDXN
16 D=N/(N-1)*(1-Y/
E12):WAIT 20
17 PRT "INDICE DE"
18 PRT "CONFIABILI
DAD=":RND(O,-3)
:PRT
19 S=E*SQR (1-D)
20 PRT "ERROR "
21 PRT "TIPICO DE"
22 PRT "MEDIDA ="
RND(S,-3):PRT
23 PRT "EQ.=":(P-K
)*100/CNT
24 LOAD #0
29 L=A(X)/1E3:O=FR
AC L*1E3:L=INT
L/1E3:S=FRAC L*
1E3
30 Q=INT L:RET

```

Resultados y discusión

El establecer la asignación de aciertos a las respuestas, durante la elaboración de la clave del examen, proporcionó el mayor aporte que pueda tener el método a la optimización de la calificación del test de ensayo, puesto que ya se han aportado muchas recomendaciones para lograr minimizar sus desventajas (2), pero ninguna de ellas contempla un método específico por el cual se pueda convertir la palabra escrita en números.

El número de aciertos se establece en base al tipo de respuesta esperada, por ejemplo:

Pregunta:

¿Qué tipo de reacciones puede sufrir el acetaldehído (CH₃CHO), dependiendo del tipo de enlace que presenta?

Respuesta esperada:

Adición nucleofílica y redox en C = O, aldocondensación en C-H del metilo.

Se establecen 7 aciertos para la respuesta: uno para cada una de las palabras claves: adición, nucleofílica, redox, C=O, aldocondensación, C-H, metilo.

Nótese que a la respuesta adición nucleofílica se le asignan dos aciertos, puesto que se evalúa el procedimiento de la respuesta correcta (según la tabla de especificaciones elaborada) (2) y no sólo la respuesta; lo mismo puede decirse de los demás aciertos asignados. El asignar un acierto a cada una de las palabras claves establece el método que se empleará para la calificación de esta respuesta ya que la calificación consistirá en identificar y cuantificar el número de aciertos que tenga una respuesta con relación a la clave ya establecida.

Una vez encontrado el número de aciertos obtenido por el estudiante, se coloca este valor al lado izquierdo de la respuesta. Este valor será multiplicado (durante el desarrollo del programa CALEXOP 3) por el porcentaje que hayamos ponderado (factor de calificación) para esta pregunta en particular al establecer el examen.

Como se habrá notado, la calificación en base a los aciertos, es rápida y sin posibilidad de error, ya que no se trabaja con porcentajes de ponderación como se acostumbra (números engorrosos) sino con números sencillos.

La información que se produce luego de introducir los aciertos, es en su orden:

	B	R	M	VAR
1)	85	0	15	0.124
2)	36	0	64	0.231
3)	16	0	84	0.137
4)	62	0	38	0.236
5)	35	0	65	0.226
6)	45	0	55	0.248
7)	73	0	27	0.198

1. Valor de la nota para cada examen.
2. Análisis estadístico de respuestas, por ejemplo, para el siguiente segmento obtenido en la calificación de un examen parcial.

La pregunta No. 3 la respondieron el 16o/o en forma correcta y un 84o/o incorrecta, con una varianza de 0.137. Con estos datos el Profesor del curso puede establecer (en base a la tabla de especificaciones) la conducta u objetivo vrs. contenido que no se alcanzó durante el curso pues es de analizar las causas por las que el 84o/o no contestara en forma correcta.

El conjunto de datos reportado durante este análisis permite que la evaluación sirva de retroalimentación, puesto que el profesor puede insistir en los puntos débiles posteriormente al examen.

Otro segmento del reporte de datos de este mismo examen representa el análisis estadístico del índice de confiabilidad, error típico de medida y balance del examen (eq):

INDICE DE
CONFIABILIDAD = 0.73

ERROR
TIPICO DE
MEDIDA = 2.18

EQ=-0.3090909091

I	PM	LRS	F	FA
20	15.5	28.5	6	6
30	25.5	30.5	9	15
40	35.5	40.5	12	27
50	45.5	50.5	12	39
60	55.5	60.5	7	46
70	65.5	70.5	4	50
80	75.5	80.5	3	53
MEDIA= 41.72				
MEDIANA= 41				
MODA= 41				
CUARTIL1= 29.33				
CUARTIL3= 53.14				
DES. STANDARD= 17.42				
CASOS= 54				

C. NORM.	YC
-10.5	0.1
-6.2	0.3
-1.6	0.5
2.5	1
6.9	1.7
11.2	2.7
15.6	4
19.9	5.7
24.3	7.5
28.6	9.3
33	10.9
37.4	12
41.7	12.4
46.1	12
50.4	10.9
54.8	9.3
59.1	7.5
63.5	5.7
67.8	4
72.2	2.7
76.5	1.7
80.9	1
85.2	0.5
89.6	0.3
94	0.1

X₁₂= 0.974991347

CERT= 71.63529413

El índice de confiabilidad orienta al profesor sobre lo "confiable" que puede ser el instrumento que usó como medida (0.5 malo, 0.6 a 0.8 bueno, 0.9 óptimo) (4) puesto que a partir de este dato puede hacer su análisis de ítems.

El "error típico de medida" orienta el grado de error que se puede cometer por nota de examen y el balance la inclinación de su examen hacia cuán difícil (menor que cero) o cuán fácil (mayor que cero).

El otro segmento importante, es el reporte estadístico, el cual muestra una tabla con sus estadísticas, y la curva de normalización de frecuencias con el valor de chiquadrada y su porcentaje de certeza. Estos datos ayudarán para ejecutar pruebas de hipótesis de medias u otras.

Cualquier información más amplia sobre el tema, consultar con el autor.

Referencias

1. Goring PA Manual de Mediciones y Evaluación del Rendimiento de los Estudios. Buenos Aires: I Kapeluz, 1978 230 p (pg. 119,128).
2. Lemus LA Evaluación del Rendimiento Escolar Buenos Aires: Kapeluz, 1974 356p (p110, 124)
3. DOWNIE NM, Heath RW Métodos Estadísticos Aplicados. México: Harla, 1973 340p (264,273).
4. Magnusson D, Teoría de los Test. México: Trillas, 1977 320p (p 38,143)
5. Cerezo Ruiz A, Estadística Descriptiva e Introducción al Análisis. Guatemala: Universitaria (col. aula), 1978 241p (81-157).



DISTRIBUIDORA DE LABORATORIO Y EQUIPO INSTITUCIONAL

SOCIEDAD ANONIMA

12 CALLE 3-31, ZONA 1

TELEFONO: 53-83-83

GUATEMALA, C. A.

APARTADO POSTAL 2925

DISTRIBUIDORES DE:

BBL	Medios de cultivo, discos de sensibilidad, sistema gaspak, sistema de identificación minitek, tiempo de coagulación fiore-sistem.	FORMA	Congeladores, Refrigeradoras, Incubadoras CO ₂
BREWER	Reactivo RPR, agitadoras para pruebas en tarjeta.	GILFORD	Espectrofotometro Stasar III, sistemas de Química Clínica.
BROOKFIELD	Viscosímetros.	KIMBLE	Cristalería para Laboratorios.
CORNING	Potenciómetros, Fotómetros de llama. Electroforesis, RIA.	SARTORIUS	Filtración en Microbiología e Industrial, Balanzas.
EALING	Equipo experimental para Biociencias, Óptica, etc.	THERMOLYNE	Agitadores, Agitadores/Calentadores, Horno/Incubadora, Vortex, Agitadores de Tubos.
		WIENER	Reactivos para Química Clínica, Reactivos para contadores Hematológicos.