

## DESARROLLO Y VALIDACION DE UNA PRUEBA PARA MEDIR CONOCIMIENTOS MINIMOS DE QUIMICA ORGANICA PARA XI AÑO DE LA EDUCACION DIVERSIFICADA

*María del Rosario Sibaja Ballesteros  
Ana Rocío Madrigal Gutiérrez  
Juan Bolaños Montero  
Rocío Pereira Esteban*

### I. Introducción

Desde 1979 se desarrolla en el Departamento de Química de la Facultad de Ciencias Exactas y Naturales de la Universidad Nacional, un programa de extensión que tiene como fin último, según lo planteado en el Modelo de Fundamentación Teórica, "Mejorar la enseñanza de la Química, y a través de esta, promover una formación científica adecuada en los futuros ciudadanos del país" (Ruiz, R., 1983, p. 2). Dentro de este programa denominado Subprograma de Mejoramiento para la Enseñanza de la Química (SUMEQ), se ejecutan una serie de acciones, a saber: readecuación de programas de X y XI año, edición de textos y elaboración de guías, capacitación, seguimiento, evaluación formativa e investigación evaluativa, (Ruiz, R., 1984) las cuales se sustentan en una fundamentación teórica (Ruiz, R. 1983) - psicológica (Carvajal, Z; Fuentes, C; Ruiz, R., 1983) epistemológica y filosófica (Meoño, R., 1983) y metodológica (Ruiz, R. 1984) que actúa como eje motor e integrador de todas ellas.

La investigación evaluativa tal y como es concebida por los miembros del equipo del SUMEQ, concepción que coincide con los planteamientos de Shaeffer y Niquiyangi (1983) quienes indican "... existe una modalidad investigativa enfocada al planeamiento para evaluar sistemas educativos o programas educativos, y diagnosticar sus problemas con el propósito de diseñar, planear y programar, hacer pronósticos y formular metas, objetivos y estrategias para la toma de decisiones y acciones" (p. 9), es la acción que permite conjuntamente con la evaluación formativa (Pereira, R. et al, 1983), analizar los logros alcanzados por estudiantes, profesores y extensionistas. Estas acciones permiten realizar

un proceso de retroalimentación para mejorar sobre la marcha el trabajo que se ejecuta. Guba plantea que "el propósito de la evaluación formativa es el refinamiento y el mejoramiento, lo que hace pensar en los aspectos intrínsecos de la entidad que está siendo evaluada" (Guba, E; Lincoln I., 1981; p. 220).

Uno de los proyectos de investigación evaluativa que se ejecuta es el denominado "INFLUENCIA DE LA METODOLOGIA PARTICIPATIVA EN EL APROVECHAMIENTO Y NIVEL DE OLVIDO DE LOS CONOCIMIENTOS DE QUIMICA ADQUIRIDOS EN LA EDUCACION DIVERSIFICADA".

Dentro de este proyecto se desarrollan actividades específicas para el programa de Química de X año (Química General) y para el de Química de XI año (Química Orgánica), actividades tendientes a la determinación del aprovechamiento de los cursos de Química recibidos, a la medición de la perdurabilidad de los conocimientos adquiridos, así como también a la evaluación global de las acciones propuestas por el SUMEQ.

Como un primer paso para el logro de los objetivos propuestos en la investigación, en lo que respecta a la Química Orgánica, es necesario desarrollar instrumentos (pruebas) válidos y confiables, con los que se pueda medir los conocimientos con que ingresan al curso los estudiantes (Perfil de entrada) y con lo que concluyen el curso (Perfil de salida) y a partir de ellos establecer el aprovechamiento del curso recibido, e iniciar las mediciones de perdurabilidad. En el año 1983, en el SUMEQ se estructuró un instrumento con un objetivo similar (Sibaja, R; Sánchez, V. 1985, p. 15). Sin embargo esta prueba no puede ser utilizada en la presente investigación, ya que los contenidos que incluye varían considerablemente con los propuestos actualmente por el SUMEQ y por el Ministerio de Educación Pública.

Previo a la elaboración de las pruebas, es necesario definir el modelo de evaluación (Normas o Criterios) bajo el cual se enmarcarán, aunque en realidad la diferencia entre estos modelos "...no está tanto en cómo se construyen las pruebas, sino en la especificidad o estrechez del dominio que cubren y la forma completa como se determinan los niveles de rendimiento de ese dominio" (Lehmann, I; Mehrens, W. 1977; p. 167).

Se decidió utilizar el modelo de evaluación con Referencia a Normas ya que "este se emplea para determinar el rendimiento de un individuo en relación con el rendimiento de otros individuos", (Popham, W; Husek, T. 1977, p. 85) y en este caso lo que interesa es el conocimiento global que sobre la disciplina, Química Orgánica, tiene una muestra de estudiantes del XI año del Ciclo Diversificado y el rendimiento alcanzado por el grupo seleccionado.

Explícitamente los objetivos propuestos para esta primera parte de la investigación son:

1. Desarrollar dos pruebas paralelas que posibiliten medir conocimientos mínimos de Química Orgánica para la Educación Diversificada.
2. Determinar la validez y la confiabilidad de las pruebas.

## II. Metodología

Para el desarrollo de las pruebas, se consultó directamente con los profesores implicados en el proceso de enseñanza-aprendizaje, tanto de Educación Diversificada como de Enseñanza Superior, en el área de la Química Orgánica, procurando obtener como resultado un panorama que reflejara la realidad del aula. El procedimiento seguido para estructurar las pruebas se desglosa a continuación.

### 1. *Determinación de los objetivos mínimos por medir*

A partir de los programas de estudio propuestos por el Ministerio de Educación Pública y por el SUMEQ, para el curso de Química Opcional de undécimo año, se estructuró una lista de

110 objetivos, dentro de los cuales se contemplan todos los contenidos que deben haber sido desarrollados al concluir el curso lectivo. La lista en cuestión fue sometida a análisis por parte de una muestra aleatoria de los profesores de Química del 20% de las instituciones de Enseñanza Media de cada una de las Regiones Educativas del país, y de profesores de las distintas instituciones de Educación Superior, incluyendo tanto los de las Sedes Centrales como los de Centros Regionales.

Se les solicitó específicamente a los docentes, que expresaran su opinión en cuanto a cuáles de los 110 objetivos del listado inicial, constituyen lo mínimo que debe dominar el estudiantado al concluir su undécimo año. Un 64% de los 75 profesores consultados dieron respuesta y prestaron colaboración; el resto no contestó.

Los criterios utilizados para la selección de estos objetivos considerados como mínimos, fueron:

1. Se elaboró un primer acercamiento con aquellos objetivos marcados por el 60% o más de los profesores consultados.
2. Para una segunda selección, se decidió eliminar de la lista aquellos objetivos que son desarrollados en décimo año, y que por razones de orden en los contenidos se repiten en XII año.
3. Por último, aquellos objetivos que estaban incluidos en otros y podían ser medidos conjuntamente, fueron eliminados de la lista final.

A partir de los resultados obtenidos del análisis anterior, se estableció un listado de OBJETIVOS MINIMOS, los cuales nuevamente fueron sometidos al análisis de cinco jueces, con la intención de ratificarlos como mínimos, para ser luego adoptados como base para la estructuración de las pruebas.

### 2. *Elaboración de preguntas de selección*

Una vez definidos los objetivos, se volvió a solicitar la colaboración de los profesores tanto de Enseñanza Media como Superior, para la elaboración de preguntas de selección única que permitieran medir los objetivos mínimos. Se obtuvo un total de 110 preguntas a las cuales debía determinárseles la validez.

### 3. *Determinación de la validez de contenido*

Para las preguntas de rendimiento académico, basadas en normas, el tipo de validez más apropiado es el denominado Validez de Contenido (Groulund, N., 1973) y para este caso en particular se siguieron los dos mecanismos más apropiados para determinar este tipo de validez, a saber: validez por criterio de jueces y cuadro de balanceo y tabla de especificaciones.

Se solicitó a siete profesionales de la enseñanza de la Química o a químicos orgánicos, que determinarían la congruencia entre cada uno de los objetivos y las preguntas propuestas para medirlos. A partir de esta valoración se eliminaron o reestructuraron algunas de ellas. Se tomó como criterio de selección la concordancia de cinco jueces como mínimo para aceptar la congruencia entre objetivo y preguntas.

Una vez seleccionadas las preguntas que representaban adecuadamente el campo específico a que se mediría, se procedió a estructurar dos pruebas paralelas tomando como base el nivel de dificultad de cada uno de los objetivos, es decir, el nivel cognoscitivo en el cual puede ubicarse según la taxonomía propuesta por Bloom (Bloom, B. 1972), por lo que el número de preguntas asociado a cada objetivo varía entre una y tres según corresponda a los tres niveles del dominio cognoscitivo con los que se trabajó, a saber, conocimiento, comprensión y aplicación.

Cuando las pruebas estuvieron estructuradas, nuevamente se sometieron al criterio de cinco jueces para ratificar su validez de contenido, así como la adecuación del balance establecido de acuerdo con el nivel de dificultad de los objetivos. Nuevamente se tomó como criterio la concordancia de cinco jueces.

### 4. *Determinación de la confiabilidad*

El índice de confiabilidad se determinó mediante formas equivalentes o paralelas de la prueba (Merhens, W; Lehman, I., 1977), denominadas como las pruebas A y B. Estas fueron aplicadas a un grupo piloto de estudiantes pertenecientes a las zonas definidas como Urbano Central y Central Periférica (Madrigan, R; Bolaños, J., 1988). Las instituciones Educativas de cada zona fueron seleccionadas aleatoriamente considerando si los profesores que imparten el curso estaban adscritos al SUMEQ o no. Por último, los alumnos que debían resolverlas fu-

eron seleccionados también aleatoriamente dentro de cada institución. Las pruebas A y B se aplicaron con un día de diferencia una de la otra, procurando así evitar que entre las dos aplicaciones se produjera nuevo aprendizaje producto del desarrollo de contenidos por parte de los profesores durante sus lecciones, lo que por supuesto afectaría los resultados obtenidos. No fue posible aplicarlas durante el transcurso del mismo día ya que la organización interna de las instituciones educativas lo hace prácticamente imposible, lo que dio la oportunidad de que los estudiantes pudieran analizar algún concepto que durante la primera aplicación no tuvieran suficientemente claro.

Con los resultados obtenidos en las dos aplicaciones se procedió a hacer un análisis de correlación Producto-Momento de Pearson para determinar el estimado de confiabilidad; además se calculó la *t* de Student para valorar el nivel de significancia de este índice.

Por último, y para determinar si las pruebas eran paralelas, se realizó análisis de varianza y diferencia de Medias (Steel, R; Torrie, J., 1973).

## III. Resultados y discusión

1. Como resultado de la evaluación de los 110 objetivos propuestos para determinar el campo que mediría la prueba, se definió un total de 41 objetivos como el MINIMO que un estudiante debe dominar al concluir el undécimo año de la Educación Diversificada.

A partir de estos objetivos se elaboraron 120 preguntas de selección, denominadas preguntas tipo; es decir, preguntas de las cuales pueden derivarse otras, que continuarán siendo válidas, ya que la estructura y el nivel de dificultad se mantienen; lo que cambia en la mayoría de los casos, son los compuestos químicos a los que se hace referencia. Esto permite estructurar continuamente otras pruebas paralelas, que pueden ser utilizadas durante el desarrollo de la investigación general, disminuyéndose así el aprendizaje de las preguntas y permitiendo entonces, medir realmente el contenido implícito en ellas.

Las 120 preguntas diseñadas fueron sometidas a un proceso de valoración para determinar su validez de contenido. Se consideró pregunta válida aquella que obtuvo como mínimo 75% de aceptación por parte de los jueces consultados. Al concluir el proceso quedó un total de 110 pregun-

tas, a partir de las cuales se estructuraron las formas A y B de la prueba, con un total de 50 preguntas cada una.

2. Las formas paralelas de la prueba fueron aplicadas con una diferencia de un día tal y como se plantea en la metodología, y se obtuvo para la estimación del índice de confiabilidad el siguiente resultado:

$$r = 0,62$$

$$t_c = 24,74 \text{ (con 31 grados de libertad)}$$

$$P(t) = 24,74 < 0,001$$

Dado que el valor obtenido para  $t$ , con 31 grados de libertad, indica que el valor obtenido para  $r$  es altamente significativo a un nivel de 0,001, es posible afirmar que las formas A y B para medir conocimientos de Química Orgánica son confiables por equivalencia (Merhns, W; Lehman, I; 1977).

Este método para estimar confiabilidad fue seleccionado debido a la necesidad de contar con pruebas paralelas, procedimiento que requiere mayores esfuerzos económicos y humanos, pero que minimiza el aprendizaje de las preguntas evitando así enseñar a resolver las pruebas, limitación que se representaría si se utilizara el método de estimación por consistencia interna (Merhns, W; Lehmann I; 1977).

La elaboración de instrumentos de medición siguiendo el procedimiento descrito anteriormente, fue necesaria ya que el subprograma de mejoramiento de la enseñanza de la Química no cuenta con instrumentos por medio de los cuales se pueda determinar la retención (o nivel de olvido) de los conocimientos de Química adquiridos en XI año, actividad que requiere repetidas aplicaciones de las pruebas para determinar los conocimientos que perduran en los estudiantes seleccionados como muestra.

Una vez estimada la confiabilidad de las pruebas, además de la igualdad de contenidos que incluye igualdad en la tabla de especificaciones, el mismo nivel de dificultad y tipo de preguntas -de selección en este caso-, es importante asegurarse de que las dos medidas sean equivalentes desde un punto de vista estadístico (Merhns, W; Lehmann, I., 1977). Para la determinación de esta equivalencia estadística se utilizaron los siguientes valores:

#### Estimadores de tendencia central y dispersión

		No. de casos	Media	Desviación Estándar
A	1era. aplicación	33	15,36	3,78
B	2da aplicación	33	18,45	4,51

#### Prueba de hipótesis para diferencia de medias (considerando varianzas iguales)

Valor de F	Prueba 2-colas	Valor de T	Grado de libertad	Prueba de dos colas
1,45	0,32	-3,02	64	0,004

Como se observa en la prueba de hipótesis para diferencia de medias, el valor de  $t$  obtenido es de -3,02 con 64 grados de libertad, en una prueba para dos colas, partiendo del hecho de que las varianzas en ambos grupos no son significativamente diferentes. Este valor calculado de  $t$  de Student es significativo a un nivel de  $\alpha = 0,004$ , con lo que se rechaza la hipótesis nula de igualdad en las medias de ambos grupos. Con esta información se deduce que efectivamente el promedio obtenido en el grupo B es más alto que el obtenido en el grupo A, pero que la variación interna de los puntajes en cada una de las dos pruebas ha demostrado ser la misma.

Es decir, ha habido un desplazamiento de la tendencia central, en la segunda prueba con respecto a la primera, y que se podría interpretar como un aprendizaje ocurrido con la administración de la primera prueba; no obstante, este aprendizaje fue consistente en todos los individuos puesto que se mantuvo la misma razón de variabilidad.

#### IV. Conclusiones

Teóricamente la estimación de la confiabilidad por equivalencia requiere que las formas paralelas de las pruebas sean aplicadas el mismo día (Merhns, W; Lehmann, I, 1977); sin embargo debido al funcionamiento interno de las instituciones de Enseñanza Media, no fue posible hacerlo así, sino que tal y como se describe en la metodología, las formas A y B se aplicaron con una diferencia de un día una de la otra, lo que produce un error implícito, producto básicamente,

del cambio en su ejecución. Pese a esto, por los resultados obtenidos puede concluirse que las formas paralelas desarrolladas para medir conocimientos mínimos de Química Orgánica son válidas y confiables.

Por otra parte, para el SUMEQ es importante contar con instrumentos de medición que le permitan evaluar las acciones que realiza, y a partir de aquí desarrollar e implementar otras acciones en procura del mejoramiento de la Enseñanza de la Química en nuestro país. El desarrollo de los instrumentos descritos en el presente trabajo, constituyen un primer intento para colaborar en este sentido.

### Bibliografía

- Bloom, Benjamín, y colaboradores. *Taxonomía de los objetivos Educativos*. Buenos Aires, Editorial Ateneo; 1972. (Traducción de Marcelo Pérez Rivas).
- Carvajal, Z; Fuentes, C; Ruiz, R. *Fundamentación Psicológica: SUMEQ*. Universidad Nacional; 1983.
- Groulund, Norman. *Medición y Evaluación en la Enseñanza*. México, Editorial Parx; 1973.
- Guba E., Lincoln I. *Effective Evaluation Improving the usefulness of evaluation results through responsive and naturalistic approaches*. San Francisco, California, Jossey-Bass Publishers; 1981.
- Lehmann, W; Mehrens, I. *Medición y Evaluación en Educación y Psicología*, Barcelona, Editorial CEAC; 1977.
- Madrigal, R.; Bolaños, J. *Marco Conceptual y Diseño de la Muestra*, Guatemala, Revista Estudio, Volumen IX, Número 4/88; 1988.
- Mehrens W., Lehmann I. *Biblioteca de Psicología Educativa*. Tomo 1, México. CECSA; 1977.
- Meoño, Rodolfo. *Fundamentación Filosófica y Epistemológica*. SUMEQ. Universidad Nacional; 1983.
- Pereira, R.; Fonseca M; Ruiz R.; *Modelo de Evaluación Formativa. Subprograma de Mejoramiento de la Enseñanza de la Química*. Resumen de ponencias, II Encuentro Nacional de Investigadores en Educación. Universidad de Costa Rica; 1983.
- Pophan, W; Husek, T. *Implications of criterion referenced measurement*, Journal of Education Measurement; 1977.
- Ruiz, Rose Marie. *Subprograma de Mejoramiento en la Enseñanza de la Química, Modelo Metodológico*. Revista Educación, Número 8, p. 119-125; 1984.
- Ruiz, Rose Marie. *Modelo de Fundamentación teórica. SUMEQ*. Universidad Nacional; 1983.
- Ruiz, Rose Marie. *Subprograma de Mejoramiento de Enseñanza de la Química, Modelo Educativo*. Revista Educación, Número 8, p. 111-117; 1984.
- Shaeffer S.; Nikiyangi J. *The environment of educational research*. International Center of research development, Ottawa Canadá, 1983.
- Sibaja, R; Sánchez, V. *Desarrollo y Validación de una Prueba de conocimientos mínimos de Química para XI año*. Resumen de ponencias, III Encuentro Nacional de Investigadores en Educación, Universidad de Costa Rica; 1985.
- Steel, R; Torrie, J. *Principles and Procedures of Statistics*, New York Mc Graw-Hill; 1973.