

Evaluación metodológica de la investigación contemporánea sobre respuestas académicas complejas en la instrucción Universitaria: Primera parte

Juan José Sánchez Sosa

Universidad Nacional Autónoma de México

ABSTRACT

A review was made of a number of research articles on the effects of diverse variables on the establishment or maintenance of "complex" academic responses. Each paper was reviewed concerning the methodological quality of its research and the experimental and educational relevance of its results. The sources of the initial survey included the standard indexes on educational publications. The methodological criteria included, among others: validity of dependent and independent variables; subjects' comparability within group comparisons; control of experimenters and observers biases; observers reliability and control of other "internal" and "idiosyncratic" confoundings. Within this first part, the following responses were explored and evaluated: originality, creative thinking, comprehension, abstraction and generalization. The conclusion of the present paper is that, in general, there seems to be some support for the inclusion of some procedures in higher education in order to foster various complex skills. However, most of the research results should be taken very carefully and tentatively because of their numerous methodological flaws.

RESUMEN

Se revisaron aquellos artículos de investigación que declaraban estudiar los efectos de diversas variables sobre el establecimiento o mantenimiento de respuestas académicas complejas en escenarios educativos universitarios o profesionales. Cada artículo se revisó respecto de la calidad metodológica de la investigación y de la relevancia experimental y

¹ El autor agradece los comentarios de los doctores Bill L. Hopkins, Laura Hernández Guzmán y Donald M. Baer, durante la preparación de este manuscrito. Durante el estudio referido en este trabajo, el autor estuvo financiado por el Consejo Nacional de Ciencia y Tecnología (CONACYT) y por la Facultad de Psicología, de la Universidad Nacional Autónoma de México (UNAM).

educativa de los resultados. Las fuentes utilizadas para la exploración inicial fueron los índices y guías estándar que cubren más de sesenta revistas y publicaciones periódicas sobre investigación educativa en varios países. Entre otros criterios de calidad metodológica se incluyeron: validez de variables dependientes e independientes; comparabilidad de sujetos en diseños experimentales de grupo; control de contaminantes derivadas de experimentadores, registradores e instructores; confiabilidad de registro o calificación; control de otros contaminantes internos y control de algunos contaminantes idiosincrásicos. Durante esta primera parte se evaluaron las respuestas de "originalidad", "pensamiento creativo", "compresión", "abstracción" y generalización. La conclusión sugiere que, en general, si bien parece haber apoyo para algunos procedimientos y materiales, la mayoría de los resultados de las investigaciones deben tomarse con muchas reservas, dada la gravedad de sus deficiencias metodológicas.

Educadores, psicólogos educativos y el hombre común y corriente, están de acuerdo en que la educación (y por tanto la instrucción universitaria), no sólo debe enseñar hechos, definiciones y el conocimiento de especificidades, sino también una constelación de repertorios académicos adicionales y/o alternativos (Gagné, 1965). Los investigadores han llamado a estas habilidades académicas "más complejas": creatividad (Maltzman, Simon, Raskin y Licht, 1960), aprendizaje de principios, solución de problemas y conceptualización (Gagné, 1964), aplicación, análisis, síntesis y evaluación (Bloom, 1965) u originalidad (Maltzman, 1960).

La mayoría de estos conceptos son derivados o aplicaciones educativas de los temas centrales en el campo de la psicología de la cognición, como pensamiento (Bruner, 1956; Bourne, Ekstrand y Dominowski, 1971), solución de problemas (Newell, Shaw y Simon, 1958), aprendizaje conceptual (Hunt, 1962) y pensamiento creativo (Wallas, 1921; Osborn, 1957). En general, se dice que estas habilidades pertenecen al campo general de las estrategias cognoscitivas de orden superior (Reitman, 1965). En términos de análisis de la conducta, se ha llamado a estas habilidades académicas: conducta conceptual (Whaley y Malott, 1971; Skinner, 1974), control de estímulos abstracto (Ferster y Perrott, 1958) o rendimiento de generalización (Miller y Weaver, 1975; Sánchez Sosa, Semb y Spencer, 1975). Cualquiera que sea la aproximación teórica que se enfoque, se considera que estas habilidades constituyen objetivos importantes del proceso de enseñanza (Gagné, 1965; Keller, 1966; Sherman, 1974).

En la última década se han investigado una amplia variedad de procedimientos diseñados para facilitar el entrenamiento de estas estrategias de orden superior en el contexto de la enseñanza a nivel universitario y profesional. En prácticamente todos los casos, los autores de estos trabajos de investigación han implicado o declarado recomendaciones derivadas de sus hallazgos sobre la mejor forma de promover dicho rendimiento académico a nivel superior.

Es claro que la razón más importante para justificar la instrumentación de un procedimiento y/o la confección de los materiales de un curso que

tenga como objeto lograr que los estudiantes adquieran algunas habilidades cognoscitivas superiores, es que la estrategia en cuestión demuestre ser realmente funcional y efectiva.

La funcionalidad de una recomendación derivada de la investigación depende de la demostración empírica de su eficacia. En gran parte, tal demostración dependerá de lo apropiado del diseño experimental utilizado (Campbell y Stanley, 1963; Sidman, 1960; Platt, 1964).

El propósito del presente trabajo es revisar la literatura de investigación contemporánea, sobre la promoción de aquellas habilidades conocidas como cognoscitivas superiores, en la instrucción profesional y universitaria, a la luz de la calidad de su metodología. La meta principal es establecer una serie de lineamientos tentativos que puedan resultar útiles fundamentalmente a dos audiencias: a aquellos educadores que piensan que las recomendaciones basadas en la investigación son valiosas y a los investigadores del comportamiento que planean investigación sobre rendimiento académico "complejo" a nivel superior.

Criterios y fuentes de evaluación

La literatura aquí revisada surgió de las fuentes y publicaciones cubiertas en los siguientes índices: *The Education Index*, desde 1966 a la fecha; *The Current Index of Journals in Education*, durante el mismo periodo; *The Annual Review of Psychology*, durante el mismo periodo; el *Journal of Applied Behavior Analysis* y otras publicaciones de orientación conductual; y por medio de la exploración de artículos relevantes a nuestro propósito y a los cuales se les podía referir a partir de las listas de referencias de otras publicaciones. En este último caso, no hubo restricciones de año; los artículos se evaluaban si resultaban relevantes.

Se hicieron las siguientes consideraciones para la inclusión de los artículos en el análisis: 1) que los artículos estuvieran listados bajo los subtítulos alfabéticos de: cognición, aprendizaje conceptual, pensamiento creativo, escritura creativa, generalización, lógica, aprendizaje de principios, solución de problemas, cuestionamiento, razonamiento y pensamiento; 2) que en la investigación descrita se hubiera usado como sujetos a estudiantes de nivel profesional o universitario; 3) que la investigación descrita reportara la medición de conducta que implicase la adquisición de habilidades, conceptos, principios, etc., relevantes a propósitos académicos, o conductas que tuvieran semejanza con habilidades académicas aunque se hubieran medido en escenarios controlados o relativamente no educativos (de laboratorio).

Se excluyeron aquellos estudios que utilizaron exclusivamente variables dependientes de tipo psicométrico (véase, por ejemplo, Anastasi, 1961) que no tuvieran una relación directa aparente con la instrucción (véase por ejemplo, Watson y Glaser, 1952). Algunos ejemplos de reactivos pertenecientes a variables dependientes de ese tipo son: "haga una lista de

todos los usos posibles que se le ocurre pueda tener un gancho de ropa de alambre”, o “imagínese que, de repente, la estatura de los habitantes del mundo se redujera a 40 centímetros. Haga una lista de todas las posibles consecuencias que podrían resultar de esto” (Renzulli, Owen y Callahan, 1974). Otros estudios de Scheffler (1965), Strope y Braswell (1966) y Henkel (1967), son buenos ejemplos de estudios excluidos por las razones mencionadas.

Organización

Los estudios que se revisaron se agruparon, primero, en base a sus variables dependientes, por ejemplo “originalidad”, “pensamiento creativo”, “comprensión”, “abstracción”, “generalización”, “pensamiento crítico”, evaluación y solución de problemas. Después, dentro de cada rubro de medida dependiente, los estudios se agruparon de acuerdo a los procedimientos utilizados para promover dichas habilidades. Se hicieron comentarios sobre los esfuerzos de cada autor para validar sus variables independientes y dependientes, controlar las contaminaciones de instructores y experimentadores, controlar o monitorear las contaminaciones de observadores y calificadores, garantizar la equivalencia inicial de los sujetos experimentales en términos de la variable dependiente y registrar la confiabilidad entre calificadores o registradores. Cuando resultaba apropiado, se hacían comentarios sobre otros puntos metodológicos relacionados directamente con el diseño experimental, y sobre resultados estadísticamente significativos que son educativamente insignificantes.

Consideraciones metodológicas

Validez. En cuanto a la validez de las medidas dependientes, si un investigador no puede garantizar que está realmente midiendo o registrando lo que declara, la conclusividad de sus resultados será muy dudosa. Pudo haber ocurrido simplemente que los procedimientos experimentales afectaron (o no pudieron afectar) algo diferente, es decir, una medida distinta a la que indica el reporte. Podría ocurrir, por ejemplo, que algún investigador declarara que un conjunto de reactivos de examen miden la capacidad de los alumnos para dar soluciones creativas a problemas de física, pero los reactivos se hayan diseñado de tal manera que todo lo que tenía que hacer el estudiante para contestarlos correctamente fuera recordar la solución dada por un ayudante de laboratorio en una demostración realizada con anterioridad. En este caso, el reactivo probablemente estaría midiendo “memorización” y no “creatividad”, en el sentido en el que se usan típicamente estos términos.

El primer paso para lograr la validación de una medida académica es definir, tan clara y conductualmente como sea posible, la tarea o habilidad que intentamos medir. Si la medida es la aplicación de un examen, se

deberán confeccionar reactivos que se ajusten a la definición. Una vez escritos los reactivos, se puede dar esta definición y los reactivos a varios jueces *independientes* (por lo general otros colegas en el área) y se les pide que describan, para cada caso, si el reactivo tiene probabilidades de medir tal habilidad. El siguiente paso del investigador será, entonces, asegurarse de controlar o monitorear cuidadosamente cualquier condición que afecte dicha validez de los reactivos (tal como la condición del ejemplo de la solución de problemas de física).

Las mismas consideraciones se pueden aplicar al respecto de la validez de la variable independiente; por ejemplo, si un investigador desea evaluar los efectos de un cuadernillo de trabajo sobre la capacidad de los estudiantes para evaluar alguna clase de eventos, un posible método es, una vez más, reunir la certificación independiente de varios jueces calificados, sobre las características reales del cuadernillo.

Contaminación de Experimentadores e Instructores. Otro tipo de contaminantes son los derivados de la actividad del instructor o experimentador. Si los instructores saben acerca de un procedimiento del salón de clase, o esperan algún resultado específico de él, su conducta al dar conferencias puede alterarse y ajustarse a dicha "expectativa" produciendo de esta manera, efectos espurios. Esto puede ocurrir aun no deliberadamente. Lo mismo se aplica al experimentador que registra la variable dependiente o a los sujetos mismos. Por tanto, es crucial mantener a los experimentadores, observadores, sujetos, etc., tan ignorantes como sea posible de cuáles son el diseño experimental y los procedimientos experimentales del estudio (Johnson y Solso, 1971). Si esto no resulta práctico, se debe entrenar observadores adicionales que se usen para monitorear cualquier actividad del personal que pueda contaminar los efectos experimentales.

Confiabilidad. La falta de confiabilidad de observadores o calificadores es otra fuente de vulnerabilidad en los datos de la investigación educativa. Si no hay acuerdo entre dos calificadores u observadores independientes que posean definiciones conductuales o criterios de calificación para exámenes, los datos pueden resultar completamente inútiles. Entre otros problemas, la definición puede no ser lo suficientemente objetiva o el método de registro puede ser insensible a los cambios de la conducta en cuestión. Una forma de mejorar la confiabilidad consiste en revisar cuidadosamente las definiciones o los criterios, o entrenar a los calificadores u observadores dentro de un sistema de realimentación.

Selección de Sujetos. La comparabilidad o equivalencia inicial de los sujetos experimentales (cuando se usan diseños de grupo) se refiere a la pregunta ¿Se deben los efectos observados en un experimento a la *selección* diferencial y la *asignación* de sujetos a los tratamientos experimentales y no al tratamiento experimental mismo? (Campbell y Stanley, 1963).

A este problema se le puede controlar por medio de la igualación de los sujetos respecto de su rendimiento en la variable dependiente. Otra alternativa es el uso de diseños experimentales intrasujetos en los cuales, los

mismos sujetos se exponen, en momentos diferentes, a condiciones experimentales distintas.

Se puede encontrar una relación detallada de las implicaciones metodológicas del control de contaminantes en Sidman (1960); Campbell y Stanley (1973); o McGuigan (1960).

Relevancia de Resultados. Finalmente, se deberá advertir la relevancia de los resultados experimentales. Una de las herramientas más comúnmente empleadas por muchos investigadores para estimar la magnitud de los efectos experimentales, son los procedimientos de prueba de hipótesis estadísticas (véase por ejemplo, Stilson, 1966). Como se mostrará en algunos ejemplos más adelante en este trabajo, a veces se encuentra que los efectos experimentales son estadísticamente significativos; sin embargo, muchos instructores o educadores nos dirían que tales efectos son educativamente insignificantes e irrelevantes especialmente si el procedimiento en cuestión es caro o difícil de instrumentar. Algunos autores recomiendan simplemente abstenerse de utilizar tales herramientas estadísticas (Sidman, 1960). Otra alternativa puede ser el incrementar el nivel de significancia para la aceptación de diferencias experimentales hasta que éstas resulten prácticas y educativamente relevantes.

El presente trabajo, incluye algunos comentarios generales y una serie de recomendaciones tentativas, básicamente para aquellos lectores activos en la investigación educativa, interesados en el mejoramiento de sus sistemas de enseñanza.

Originalidad y “pensamiento creativo”

No existe mucho acuerdo respecto de la definición de originalidad y creatividad en general. Un aparente común denominador de muchas concepciones parece ser el criterio de novedad, es decir la “novedad” o lo poco usual de las respuestas de los sujetos a reactivos de examen, así como un número de respuestas no duplicadas.

Tres estudios principales han explorado los efectos de diferentes procedimientos sobre la promoción de originalidad o creatividad en el rendimiento académico de estudiantes de nivel profesional o universitario. En un estudio hecho por Frederiksen y Evans (1974) se asignó a 395 estudiantes a un grupo control y dos grupos experimentales. Los sujetos de uno de los grupos experimentales recibieron una lista larga de hipótesis de investigación “aceptables” relativas a una serie de datos experimentales. Los sujetos del segundo grupo experimental recibieron una lista corta conteniendo “sólo las mejores” hipótesis relativas a los mismos datos. Los sujetos del grupo control simplemente llenaron un cuestionario no relacionado a los datos de investigación. Los resultados mostraron que el rendimiento de los sujetos de los grupos experimentales fue ligeramente superior al de los sujetos del grupo control respecto a la originalidad (número total de hipótesis no duplicadas). Asimismo, las hipótesis dadas por los

sujetos a los que se les dio la lista larga fueron más numerosas pero de menor calidad que las hipótesis dadas por los sujetos a los que se les había dado la lista corta. Los autores mencionan, pero no describen, un tipo de evaluación estadística de la validez de la variable dependiente (consistencia interna). No se reportan en el artículo controles para ninguna de las restantes fuentes de contaminación discutidas con anterioridad.

En otro estudio, McCormack (1971) asignó al azar a 69 estudiantes de un curso sobre introducción al método científico, a un grupo control y uno experimental. A los sujetos del grupo control se les enseñó por medio de un formato tradicional de conferencia, mientras que el grupo experimental se exponía a: "discusión espontánea" (brainstorming), "desarrollo de cuestionamiento" (inducir a los estudiantes a exigir datos que apoyaran algún resultado), "análisis morfológico de problemas" (para aquellas relaciones que podían resultar en explicaciones originales), e "invitaciones al pensamiento creativo". Los sujetos experimentales, de acuerdo a los autores, rindieron significativamente mejor que los sujetos control en la contestación de preguntas de examen que requerían fluidez (número total posible de atributos de un evento), flexibilidad (mención de aplicaciones poco comunes de algún objeto), y originalidad (novedad de explicaciones para un evento). En este artículo no se muestran datos de ningún tipo, por tanto no es posible estimar la talla de las diferencias descritas y no se reporta control sobre ninguno de los contaminantes discutidos, a menos que se considere a la asignación al azar como un control adecuado de la contaminación de selección de sujetos.

En un estudio más reciente de Johnson (1975), se igualó a 75 estudiantes de psicología respecto de su rendimiento en un examen parcial y se les asignó a un grupo control y a uno experimental. Todos los sujetos tenían que entregar tres trabajos bimestrales, y se les regresaban evaluados y con comentarios escritos en el margen por el instructor. Los trabajos del grupo experimental tenían comentarios como: "¿podrías decir esto con tus propias palabras?", "trata de usar ejemplos de tu propia experiencia", "se agradecerán las ideas originales propias que puedas insertar". El instructor también elogiaba cualquier viso de originalidad (no se da definición) en los trabajos del grupo experimental. A estos sujetos también se les daban prácticas sobre problemas no académicos e instrucciones para que fueran "originales". El grupo control únicamente recibió comentarios neutros a sus trabajos. Las calificaciones de originalidad en los exámenes parcial y final del grupo experimental fueron aproximadamente 30 por ciento más altas que aquellas de los sujetos control; sin embargo, los autores no explican un enorme traslapamiento entre la variabilidad de las respuestas de ambos grupos. Un coeficiente de correlación entre los puntos de originalidad entre el primer calificador y el segundo (no se reporta independencia entre ellos) dio por resultado una $r = 0.73$. El autor también reporta control sobre la contaminación del experimentador dejando transcurrir un periodo de dos semanas entre la asignación de sujetos experimentales y controles y las res-

puestas del experimentador al rendimiento de los sujetos. Se podría decir, por tanto, que este estudio controló la incomparabilidad inicial de los sujetos y las contaminaciones del experimentador y contó con una medida de confiabilidad, aunque el control de contaminación de experimentador es de dudosa eficacia. No se reportan procedimientos para asegurar la validez de la variable independiente ni de la dependiente, ni el control de la contaminación de calificadores. El gran traslapamiento de los puntos de ambos grupos resta a estos resultados una cantidad considerable de importancia.

De esta manera, la investigación en esta área incluye un apoyo bastante cuestionable de la eficacia de modelos de rendimiento creativo, la discusión libre, el desarrollo de cuestionamiento y el análisis morfológico de problemas, sobre la promoción de originalidad y pensamiento creativo; sin embargo, se observa un apoyo un tanto más sólido para el uso de la realimentación y las instrucciones complementarias. Sin embargo, estos resultados se deben tomar con algunas reservas, dados los graves problemas metodológicos descritos.

“Compresión”, “abstracción” y generalización

También en esta área existe muy poco acuerdo entre investigadores respecto de las definiciones de estas habilidades académicas; a veces se refiere al fenómeno conductual conocido como “transferencia” (Gagné, 1965).

Sin embargo, con base en algunas consistencias en el uso de varias medidas dependientes, se vislumbra una concordancia general. Estas habilidades se refieren, en general, a la emisión de algunas respuestas que no se aprenden “directamente” pero que en algún aspecto son similares a otras que sí se han aprendido directamente (Gagné, 1965). Se dice, por ejemplo, que la generalización ocurre cuando aparecen respuestas nuevas dentro del repertorio académico de los estudiantes, siempre y cuando estas respuestas no se hayan modelado previamente o se hayan entrenado directamente; por ejemplo, un estudiante es capaz de identificar correctamente la relación entre una clase de estímulos verbales y una instancia o ejemplo que la ilustre, no habiéndose enseñado tal ejemplo durante el curso (Sánchez Sosa, con base en Lutzker y Sherman, 1974; Whaley y Malott, 1971).

Un estudio investigó los efectos de dos procedimientos de enseñanza sobre el aprendizaje de principios (Cole y Raven, 1969). Se asignó a 38 alumnos de nivel universitario a uno de los siguientes tratamientos. En un grupo, a los estudiantes se les enseñaba a analizar problemas de física por medio de instrucción sobre los principios correctos, sin mencionar principios falsos o no aplicables a dichos problemas. Otro grupo aprendió tanto los principios correctos como los irrelevantes o incorrectos. El tercer grupo sirvió como control y recibió instrucción a través de conferencias tradicionales.

Los sujetos que recibieron la instrucción referente a la aplicación de principios tanto adecuados como irrelevantes, rindió más alto que los demás sujetos, en la solución de un reactivo de examen que les pedía explicar por qué un cuerpo físico determinado podría flotar o no.

Aparte de la asignación al azar de los sujetos a los tratamientos, no se reporta esfuerzo alguno por controlar la invalidez de las variables dependientes o independientes, de las contaminaciones de observadores o instructores, y de la falta de confiabilidad de registro o calificación.

Dos estudios han investigado los efectos de métodos de enseñanza estructurados por el instructor, o bien, por el grupo de estudiantes, sobre la ejecución en preguntas de examen de "generalización conceptual" y "comprensión". Coop y Brown (1970) encontraron que las conferencias estructuradas por el profesor producían calificaciones más altas en generalización conceptual, que clases independientes de solución de problemas (las cuales no se describen en detalle). Los resultados se afectaron tanto a estudiantes que se suponía eran "analíticos" tanto como para aquellos que supuestamente eran "no analíticos", de acuerdo a una medición psicométrica previa. Los autores no definieron operacionalmente ninguno de los términos escritos entre comillas, no se reportaron esfuerzos para controlar ninguna de las variables contaminantes: lo que es más, ni siquiera los materiales y lecturas del curso eran los mismos para ambos grupos.

En un estudio anterior (Ward, 1956) se sugiere que los estudiantes "más capaces" (los que sacaron las calificaciones más altas en una prueba de inteligencia) rendían mejor en reactivos de examen diseñados para medir "comprensión" (no se da una definición) si ellos mismos decidían cuáles serían los objetivos de instrucción, las discusiones y actividades en clase y las evaluaciones. A este procedimiento se le llamó "enseñanza decidida por el grupo". Por otra parte, los estudiantes que eran "menos capaces" rendían mejor bajo condiciones de enseñanza decidida por el instructor. Los autores obtuvieron un coeficiente de confiabilidad de consistencia interna de $r = 0.77$ sobre las pruebas académicas y reportan haber asignado los sujetos a los tratamientos al azar. No se reportan ni confiabilidad de calificación ni controles de invalidez de las variables dependiente e independiente, ni el control de ningún tipo de contaminación derivada del personal del curso.

A partir de esta exploración, parece claro que los resultados en esta área son confusos y poco concluyentes. Tal vez se pueda sugerir modestamente un apoyo para la enseñanza decidida por el instructor; sin embargo, no se controlaron demasiadas contaminantes, incluyendo la crucial validación de las variables independientes.

Otros estudios se han enfocado sobre los efectos diferenciales de materiales de estudio complementarios sobre la "comprensión", "abstracción" y generalización. Dichos materiales usualmente constituyen ejercicios o preguntas de estudio. Se les llama comúnmente "organizadores" (Ausubel, 1960), preguntas auxiliares (Fraser, 1971), o preguntas de estudio (Miles, Kibler y Pettigrew, 1967). Scandura y Wells (1967) mostraron que el

rendimiento promedio de los estudiantes en la solución de problemas matemáticos sobre teoremas que requerían una definición y/o una aplicación, axiomas, y la aplicación de teoremas a situaciones nuevas, era aproximadamente 11 por ciento más alto en aquellos estudiantes a los que se les habían dado organizadores de avance (antes de estudiar), los cuales tenían el formato de juegos matemáticos.

Los estudiantes a los que se les dieron introducciones históricas irrelevantes al material presentaron un rendimiento menor en los reactivos de examen. Los resultados mostraron una diferencia menor a un reactivo de examen correcto por estudiantes; a pesar de ello, los autores declararon que las diferencias eran estadísticamente significativas. En este estudio, los estudiantes contestaron una preprueba pero no se hicieron comparaciones intrasujetos. Los autores reportaron que la dificultad de los reactivos se mantuvo constante pero no muestran un procedimiento de confiabilidad para esta consistencia. No se reportan procedimientos de validación ni precauciones para controlar contaminaciones del personal o la falta de confiabilidad de calificación.

En otro estudio (Rickards y DiVesta, 1964) un grupo de sujetos recibió un conjunto de preguntas de estudio de "aprendizaje significativo" las cuales los instruían para que organizaran los datos de oraciones subordinadas bajo las "ideas" contenidas en una oración tópica supraordenada. Estas preguntas empezaban usualmente con el formato "¿Cómo sabría usted si...?". A otro grupo de alumnos se le dieron preguntas que incluían el simple "aprendizaje directo de hechos", es decir preguntas en las que los alumnos tenían que memorizar pasajes literales del texto. Los alumnos que tuvieron las preguntas de aprendizaje significativo rindieron más alto en exámenes que contenían ambos tipos de preguntas, que aquellos que habían recibido las preguntas de tipo directo. Además, los efectos fueron más grandes cuando las preguntas de estudio eran relevantes al texto. El estudio asignó los sujetos al azar a los tratamientos experimentales y también incluyó un coeficiente de confiabilidad (r de Pearson = 0.96) entre calificadores independientes que registraron el número de ideas generadas. No se reportaron en este estudio, ni controles para contaminaciones de personal ni procedimientos de validación.

Tennyson, Woolley y Merrill (1972) estudiaron los efectos de exposición de los alumnos, asignados al azar a dos tipos de ejemplos de un ritmo específico en la métrica poética del idioma inglés. La identificación y clasificación correcta de los ejemplos de la métrica fue más alta si los ejemplos habían tenido atributos irrelevantes muy diferentes (otras características de la métrica). Los ejemplos que contenían atributos irrelevantes similares produjeron sobregeneralización en los estudiantes y por tanto clasificaciones incorrectas. Las comparaciones también incluyeron varios niveles de dificultad. La dificultad se definió como la proporción de sujetos que habían clasificado correctamente en un curso anterior; sin embargo, no se validó ninguna de las variables independientes (similitud entre los atributos irrelevantes y relevancia de los atributos).

Tampoco se validaron las variables dependientes (clasificación correcta, sobregeneralización, etc.), no se obtuvo confiabilidad de calificación ni registro sobre la ocurrencia de estas variables y no se instrumentaron controles para las contaminaciones de tipo personal.

Finalmente, se han realizado otros tres experimentos, estos últimos en el contexto de la enseñanza basada en los principios del análisis experimental de la conducta. Los tres estudios se realizaron dentro del marco de referencia del sistema de instrucción personalizada (Keller, 1968). En el primero (Semb, 1974), los resultados mostraron que un criterio alto de dominio del material (los estudiantes completaban cada examen con un 100 por ciento de respuestas correctas antes de poder proseguir con el siguiente examen), y asignaciones cortas de lectura, produjeron de un 10 a un 20 por ciento más de respuestas correctas en reactivos "sonda" (reactivos no vistos por los estudiantes previamente), que bajo una condición de dominio baja (60 por ciento) y asignaciones de lectura largas. La confiabilidad de calificación fue de 92 por ciento. El estudio no contiene reporte sobre la validación de los reactivos de examen ni menciona control de contaminaciones del instructor. El experimento usó un diseño de reversión intrasujetos en el cual los estudiantes fueron su propio control. Este diseño controló el contaminante de la falta de equivalencia inicial de los sujetos (Sidman, 1960).

En otros dos estudios (Miller y Weaver, 1974, 1975) describieron un libro de texto (Miller, 1975) que contenía un componente de entrenamiento de generalización. Cada capítulo del texto estaba compuesto de una introducción, una lista de definiciones conceptuales (Skinner, 1953), un "programa de generalización" y una prueba de autoevaluación. Por medio de la adición de cada uno de los tres componentes mencionados, a través de un diseño de línea base múltiple, los autores aislaron la contribución relativa de cada componente en la medida en que se iba añadiendo a los efectos de los componentes previos. El programa de generalización consistió en ejemplos cotidianos que requerían respuestas de completamiento. La prueba de generalización consistió en ejemplos nuevos en los que se pedía a los estudiantes que describieran el concepto que se estaba ilustrando en cada ejemplo. Los resultados mostraron un incremento de aproximadamente 40 por ciento en el rendimiento de generalización de los estudiantes después de la introducción de los tres componentes del sistema. Los estudios incluyeron una evaluación sobre la validez del programa de generalización y utilizaron tanto comparaciones de grupo como comparaciones intrasujetos para evaluar los efectos. No se reporta en los estudios ni confiabilidad de calificación, ni monitoreo de los contaminantes de los instructores; sin embargo, el diseño de línea base múltiple sí controla el contaminante de la falta de equivalencia inicial de los sujetos.

El tercer estudio (Sánchez Sosa, Semb y Spencer, 1975) mostró que si se les daba a los estudiantes una lista de preguntas de estudio durante una secuencia de secciones de un curso elegida aleatoriamente, los estudiantes rendían de 10 a 15 por ciento más alto en respuestas a preguntas de examen,

que para contestarse correctamente requerían: *a*) identificar conceptos en ejemplos nuevos; *b*) aplicar los principios del curso a la solución de problemas ficticios de la vida cotidiana, y *c*) describir relaciones lógicas (del tipo “si $A = B$, y $B = C$, entonces $A = C$ ”) entre conceptos o principios. Estas relaciones lógicas condicionales no se encontraban especificadas ni en las lecturas del curso ni a lo largo de las actividades del personal del mismo. Se estableció la validez de la variable dependiente (los reactivos de examen) por medio del consenso de cuatro instructores independientes. Hubo una confiabilidad de calificación de 88.3 por ciento. El estudio utilizó un diseño de reversión intrasujetos e incluyó el monitoreo de la interacción de los alumnos con los monitores del curso, para evitar la contaminación derivada de una posible ayuda no autorizada de los monitores a los estudiantes.

En resumen, la eficacia de diferentes materiales de estudio complementarios o del arreglo de materiales con características específicas, sobre el rendimiento de generalización o “abstracción” y “comprensión” parece estar relativamente bien apoyada en la literatura de investigación a pesar de que aún varios estudios no contienen controles muy adecuados.

En la segunda parte de esta revisión metodológica, se explorará la calidad de la investigación experimental sobre las respuestas académicas complejas de “pensamiento crítico”, “evaluación” y “solución de problemas”.

REFERENCIAS

- Anastasi, A. *Psychological Testing*. Nueva York: Macmillan Co., 1961.
- Ausubel, D.P. The use of advance organizers in the learning and retention of meaningful material. *Journal of Educational Psychology*, 1960, 51, 267-272
- Bloom, B. S. (dir.) *Taxonomy of Education Objectives Cognitive Domain*. Nueva York: Longmans, Green and Co., 1956.
- Bourne, L.E., Ekstrand, B.R. y Dominowski, R.L. *The Psychology of Thinking*. Englewood Cliffs, N. J.: Prentice-Hall, Inc. 1971.
- Bruner, J.S., Goodnow, J. y Austin, G. A. *A Study of Thinking*. Nueva York: John Wiley & Sons, 1956.
- Campbell, D.T. y Stanley, J.C. *Experimental and Quasi-Experimental Designs for Research*. Chicago: Rand McNally and Co., 1963.
- Cole, H. y Raven, R. Principle learning as a function of instruction on excluding irrelevant variables, *Journal of Research in Science Teaching*, 1969, 6, 234-241.
- Coop, R. H. y Brown, L.D. Effects of cognitive style and teaching method on categories of achievement. *Journal of Educational Psychology*, 1970, 61, 400-405.
- Ferster, C.B. y Perrott, M.C. *Behavioral Principles*. Nueva York: N.Y. Appleton-Century-Crofts, 1968.
- Frase, L.T. Effect of incentive variables and the type of adjunct question upon text learning. *Journal of Educational Psychology*, 1971, 62, 371-375.
- Frederiksen, N. y Evans, F. R. Effects of models of creative performance on ability to formulate hypotheses. *Journal of Educational Psychology*, 1974, 66, 67-82.
- Gagné, R. M. The implications of instructional objectives for learning. En C. M. Lindvall (dir.) *Defining Educational Objectives*. Pittsburgh, Penn.: University of Pittsburgh Press, 1964.
- Gagné, R. *The Conditions of Learning*. New York: N.Y.: Holt, Rinehart & Winston, 1965.
- Henkel, E. T. Undergraduate physics instruction and critical thinking ability. *Journal of Research in Science Teaching*, 1967, 5, 89-94.
- Hunt, E. B. *Concept Learning*. Nueva York: John Wiley and Sons, Inc. 1962.
- Johnson, D. M. Increasing originality on essay examinations in psychology. *Teaching of Psychology*, 1975, 2, No. 3, 99-102.

- Johnson, H.H. y Solso, R. L. *Experimental Design in Psychology: A Case Approach*. Nueva York: Harper and Row, 1971.
- Keller, F. S. A personal course in psychology. En Ulrich, R., Stachnik, T. y Marbry, J. (dirs.), *Control of Human Behavior, Vol. 1*. Glenview, Ill.: Scott, Foresman, & Co., 1966, pp. 91-93.
- Lutzker, J. y Sherman, J. A. Producing generative sentence usage by imitation and reinforcement procedures. *Journal of Applied Behavior Analysis*, 1974, 7, 447-460.
- Maltzman, I. On the training of originality. *Psychological Review*, 1960, 14, 4, 22-242.
- Maltzman, I., Simon, S., Raskin, D., y Licht, L. Experimental studies in the training of originality. *Psychological Monographs*, 1960, 74, (6, No. 493).
- McCormack, A. J. Effects of selected teaching methods on creative thinking, self-evaluation, and achievement of students enrolled in an elementary science education methods course. *Science Education*, 1971, 55, 301-307.
- McGuigan, F. J. *Experimental Psychology: A Methodological Approach*. Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1960.
- Miles, D. T., Kibler, R. J. y Pettigrew, E. L. The effects of study questions on college students' test performance. *Psychology in the Schools*, 1967, 32, 25-26.
- Miller, L. K. *Principles of Everyday Behavior*. Monterey, Calif.: Brooks/Cole, 1975.
- Miller, L. K. y Weaver, H. A behavioral technology for producing concept formation in university students. Paper read at the *Conference on Behavior Research and Technology in Higher Education*, Atlanta, Georgia, octubre, 1974.
- Miller, L. K. y Weaver, H. The use of concept programming to teach behavioral concepts to university students. En J. Johnston (dir.), *Behavior Research and Technology in Higher Education*. Springfield, Ill.: Charles C. Thomas Publishers, 1975.
- Newell, A., Shaw, J.C. y Simon, H. A. Elements of theory of human problem solving. *Psychological Review*, 1958, 65, 151-166.
- Osborn, A.F. *Applied Imagination*. Nueva York: Scribener's, 1957.
- Platt, J.R. Strong inference. *Science*, 1964, 146, 347-353.
- Reitman, W. R. *Cognition and Thought*. Nueva York: John Wiley and Sons, 1965.
- Renzulli, J.S., Owen, S.V., y Callahan, C.M. Fluency, flexibility and originality as a function of group size. *Journal of Creative Behavior*, 1974, 8, 107-113.
- Rickards, J.P. y DiVesta, F.J. Type and frequency of questions in processing textual material. *Journal of Educational Psychology*, 1974, 66, 354-362.
- Sánchez-Sosa, J.J., Semb, G. y Spencer, R. Using study guides to promote generalization performance in university instruction. Trabajo leído en la 83ª convención anual de la *American Psychological Association, Association*, Chicago, Ill., 1975.
- Scandura, J.M. y Wells, J.N. Advance organizers in learning abstract mathematics. *American Educational Research Journal*, 1967, 4, 295-301.
- Scheffler, W.C. A comparison between inductive and illustrative laboratories in collage biology. *Journal of Research in Science Teaching*, 1975, 3, 218-223.
- Semb, G. The effects of mastery criteria and assignment length on college students test performance. *Journal of Applied Behavior Analysis*, 1974, 7, 61-70.
- Sherman, G.A. PSI: Some notable failures. En G.A. Sherman (dir.) *Personalized System of Instruction, 41 Germinal Papers*. Menlo Park, Calif.: W.A. Benjamin Inc., 1974, págs. 120-124.
- Sidman, M. *Tactics of Scientific Research*. Nueva York: Basic Books, 1960.
- Skinner, B.F. *Science and Human Behavior*. Nueva York, McMillan, 1953.
- Skinner, B.F. *About Behaviorism*, Nueva York, Alfred A. Knopf, 1974.
- Stilson, D.W. *Probability and Statistics in Psychological Research and Theory*. San Francisco: Holden-Day, Inc., 1966.
- Strope, M.B. y Braswell, A.L. A comparison of factual teaching and conceptual teaching in introductory college astronomy. *Journal of Research in Science Teaching*, 1966, 4, 95-97.
- Tennyson, R.D., Woolley, F.R. y Merrill, M.D. Exemplar and non-exemplar variables which produce correct concept classification behavior and specified classification errors. *Journal of Educational Psychology*, 1972, 63, 144-152.
- Wallas, G. *The Art of Thought*. Nueva York, N.Y.: Harcourt, Brace and World, 1921.
- Ward, J.N. Goup study versus lecture demonstration method in physical science instruction for general education college students. *Journal of Experimental Education*, 1956, 24, 197-210.
- Watson, G. y Glaser, E.M. *Watson-Glaser Critical Thinking Appraisal*. Terrytown-on-Hudson, N.Y.: World Book Co., 1952.
- Whaley, D.L. y Malott, R.W. *Elementary Principles of Behavior*, Nueva York Appleton-Century-Crofts, 1971.