Universidad Nacional Autónoma de México

FACULTAD DE PSICOLOGIA

DIVISION DE ESTUDIOS DE POSGRADO

ESTIMACION EMPIRICA DE LA CAPACIDAD PREDICTIVA DE LA TECNICA DE DISEÑO LOGICO DE EXAMENES

T E S I S

QUE PARA OBTENER EL GRADO DE:

MAESTRO EN PSICOLOGIA EDUCATIVA

P R E S E N T A :

JOSE GUILLERMO SOLANO FLORES

SINODALES:

DIRECTOR DE TESIS:

Mtro. Javier Aguilar

Mtro. Miguel López Olivas
Mtra. Sandra Castañeda
Mtra. Silvia Macotela
Mtra. Gloria Marmolejo

MEXICO, D.F. 1989

M

I N D I C E

A G R A D E C I M I E N T O S

Agradezco a las personas que a continuación menciono, su influencia en este trabajo.

El maestro Javier Aguilar, asesor de esta tesis, mostró desde el principio del trabajo un interés alentador; además, aportó sugerencias e indicaciones que me ayudaron a realizar adecuadamente la investigación. La maestra Sandra Castañeda hizo comentarios tan valiosos y agudos que ampliaron mi propia percepción del trabajo y de sus posibilidades. El maestro Miguel López Olivas, siempre preocupado por el rigor metodológico, fue determinante en la realización del trabajo experimental. La maestra Gloria Marmolejo aportó comentarios y sugerencias que sirvieron para dar un importante sustento metodológico a la investigación. La maestra Silvia Macotela me ayudó a enriquecer las conclusiones y me hizo reflexionar con más detenimiento en las implicaciones de las ideas expuestas.

En distintos momentos cada uno, Octavio Torres, Daniel Zarabozo y Carlos Santoyo, emitieron interesantes opiniones sobre el análisis estadístico y los fundamentos teóricos. Ayita Ruiz-Primo, además de ser mi esposa, hizo todo tipo de observaciones críticas y me ayudó en muchos sentidos. Adrián Medina me ayudó a aplicar algunos exámenes. Rolando Díaz-Loving prestó su apoyo en las últimas fases de la investigación.

Benito Ramírez me ayudó pacientemente a manejar los archivos y los programas de cómputo. Javier Arao y Margarita Hurtado prestaron amablemente todo el apoyo de cómputo requerido. Gloria Guerrero, Columba Domínguez y Consuelo Corona hicieron la captura de los datos.

Doris Robles, Verónica Vega y Porfirio Echauri, consiguieron el acceso a algunas escuelas, en una fase preliminar de la investigación. Las profesoras Betty Feldman, Luz María Mólgora y Luz Elena Angulo, y el profesor José Ocaranza, colaborarom otorgando las facilidades necesarias para aplicar los exámenes a los alumnos en las escuelas que dirigen: Escuela Nacional Educativa, Escuela Mexicana Americana, Escuela Dos Naciones Unidas y Colegio de Montaignac. Las profesoras de los alumnos con los que se realizó la investigación colaboraron de manera entusiasta y desinteresada.

R E S U M E N

El **Diseño** Lógico de Exámenes es una técnica de naturaleza lógica que permite elaborar exámenes para evaluar el aprendizaje de procedimientos. Se basa en el empleo de algoritmos para la representación de los procedimientos y en la identificación de variables lógicas, cuya combinación define a las características de cada problema a resolver. Tres conceptos son fundamentales en el diseño lógico de exámenes: (1) ***iteratividad*** (cantidad de "vueltas" o veces que hay que aplicar un procedimiento para resolver un problema determinado); (2) ***grado múltiple*** (cantidad de trayectorias diferentes que hay que seguir al resolver ese problema); y (3) ***arreglo múltiple*** (especificación de la combinación de variables lógicas que describe sintéticamente a las trayectorias que hay que seguir para resolver el problema).

En esta investigación se pretendió determinar qué tanto la iteratividad y el grado múltiple pueden servir para predecir la dificultad de los reactivos de un examen. También se trató de demostrar que dos reactivos cualesquiera que tienen un mismo arreglo múltiple son empíricamente equivalentes: son resueltos ambos correctamente o ambos incorrectamente por una proporción significativa de alumnos. Se empleó una muestra de 211 alumnos de sexto año de primaria, que resolvieron un examen consistente en números romanos que había que escribir en números arábigos.

Se encontró que el grado múltiple no es un buen predictor de la dificultad de los reactivos. En cambio, la dificultad es una función lineal directa de la iteratividad de los reactivos. También se encontró que los reactivos que tienen un mismo arreglo múltiple son resueltos ambos correctamente o ambos incorrectamente por una proporción signifcativa de alumnos. Los resultados indican que, además de servir para elaborar exámenes de acuerdo con criterios objetivos, los criterios lógicos y formales del diseño lógico de exámenes permiten predecir la dificultad de los reactivos.

Los resultados sugieren que el DLE no solamente permite la elaboración objetiva de exámenes, sino también que la técnica puede ser útil para investigar estrategias de solución de problemas. Adicionalmente, los principios descriptivos y analíticos de la técnica pueden ser aprovechados para la secuenciación de la enseñanza.

La principal limitación de la técnica puede ser la resistencia que tanto especialistas como no especialistas pueden llegar a oponer antes de incorporarla a sus prácticas de trabajo.

INTRODUCCION

En la literatura sobre evaluación del aprendizaje, el tema de los exámenes ocupa un lugar prominente. La importancia de los exámenes en todo sistema instruccional ya ha sido señalada vehementemente por Roid (1979). No obstante, la atención de los autores se ha centrado más en la calificación y el tratamiento estadístico de los datos obtenidos al aplicar exámenes (por ejemplo: Nunnally, 1964, 1967; Payne, 1968; Gronlund, 1971; Ebel, 1972; Thorndike y Hagen 1977), que en la elaboración misma de los exámenes. Autores como Millman (1980) y Roid y Halladyna (1982) sostienen que en la actualidad la elaboración de exámenes y, sobre todo, la redacción de reactivos, se realiza de acuerdo con un conjunto de principios basados en la experiencia, pero de manera artesanal, sin una base científica.

A pesar de que existen excelentes manuales y guías para la elaboración de reactivos (véase por ejemplo: Coffman, 1971; Gronlund, 1971; Ebel, 1972; Conoley y O'Neil, 1979; Schyfter, Stockton, Trejo y García, 1979), ellos se ocupan casi únicamente de los aspectos formales de la redacción y de la manera de evitar errores que atenten contra la validez de los reactivos, pero no se ocupan de cómo determinar el contenido de un examen. Si bien se ha argumentado que los contenidos, los bancos de reactivos, los objetivos de enseñanza y el trabajo coordinado de colegas y especialistas son, entre otros, aspectos que deben ser considerados al elaborar exaámenes (véase Conoley y O'Neil, 1979), quien elabora un examen trabaja en gran medida intuitivamente y basándose en el sentido común (Bormouth, 1970). Se ha encontrado, por ejemplo, que los reactivos elaborados por dos personas difieren en contenido y calidad, aun cuando esas personas cuenten con la misma información acerca de los objetivos de enseñanza (Roid y Haladyna, 1978). Se carece, por lo tanto, de la metodología necesaria para derivar el contenido de un examen a partir de un contenido de enseñanza (Roid y Haladyna, 1982).

Como resultado de tal carencia, pueden ser identificadas tres deficiencias de las que adolece el trabajo de elaboración de exámenes. La primera es muy sencilla de enunciar: Aparte de los administrativos o prácticos, no existen suficientes fundamentos metodológicos que justifiquen el empleo de una cierta cantidad de reactivos en un examen. Este problema no ha recibido suficiente atención a pesar de haberse reconocido que la cantidad de reactivos es un factor que no debe ser ignorado al evaluar una habilidad (Priestley, 1982). De tal suerte, es difícil determinar si el aprendizaje de un tema o una habilidad es evaluado con la debida suficiencia.

La segunda deficiencia consiste en que, aparte de la metodología del análisis de reactivos (véase Acosta y Stockton, 1979), que permite, entre otras cosas, determinar estadísticamente la dificultad de un reactivo, no existen medios para agrupar a los reactivos de acuerdo con propiedades que no se relacionen con la proporción de alumnos que los resuelven equivocadamente. Considérese a los reactivos: *42384729/19929* y *-12.4378/11.4.*

La dificultad de cada uno de ellos sólo puede estimarse ***a posteriori***, según los resultados de la aplicación del examen. Sin embargo, existe un conjunto de factores que podrían ser considerados ***a priori.*** Estos factores se refieren a la ***complejidad de un reactivo***: la cantidad de características diferentes cuya combinación determina la composición del reactivo. En el ejemplo expuesto, estas características pueden ser: que el dividendo sea mayor, menor o igual que el divisor; que el dividendo y el divisor tengan signo igual o diferente; que el dividendo o el divisor, o ambos, tengan fracción decimal; o que la cantidad de dígitos del dividendo sea mayor, menor o igual que la cantidad de dígitos del divisor.

Contar con un medio para estimar la complejidad de un reactivo ofrece ventajas para quienes elaboran exámenes: por las presiones de tiempo o por la carga de carga de trabajo, no siempre es posible manipular la dificultad de los reactivos de un examen basándose en los resultados del análisis de reactivos. Cuando ello sucede, la estimación de la complejidad de reactivos puede ser un recurso complementario con el que pueda elaborarse un examen siguiendo criterios objetivos.

La tercera deficiencia se refiere al contenido mismo de los exámenes. En virtud de que la redacción de reactivos está influida por factores idiosincráticos (Roid y Haladyna, 1978), con la metodología disponible en la actualidad es imposible tener la certeza de que una persona o un equipo de personas ha cubierto en un examen los asectos curciales o esenciales de aquéllo que se pretende evaluar. En una investigación previa (Solano-Flores, 1985), se encontró que aun los especialistas de un mismo tema eran incapaces de elaborar exámenes exhaustivos que cubrieran todas las modalidades posibles de un tipo de problemas, a pesar de que se les había proporcionado instrucciones precisas para que lo hicieran. La experiencia o la formación académica de quienes elaboran exámenes puede asegurar que ciertos reactivos realmente evalúen un tema o una habilidad, pero no garantiza que tales reactivos serán los mejores o los suficientes para evaluar el tema o la habilidad.

La anterior idea es congruente con la evidencia sobre pericia en habilidades. Anderson (1985) ha señalado que los expertos en habilidades especificas utilizan una estrategia ***prodroma*** (hacia adelante) al resolver problemas del tipo en que son expertos. Esta estrategia prodroma consiste en seguir la secuencia adecuada de pasos sin considerar secuencias alternativas. En cambio, una persona inexperta sigue una estrategia ***retrodroma*** (hacia atrás): considera todas las secuencias alternativas, aun las no pertinentes, antes de resolver el problema. Según Anderson, la eficiencia del experto se debe a que identifica oportunamente las características del problema que tiene ante sí y sigue de manera automática la secuencia requerida de acciones. Plantear un conjunto de reactivos que evalúen cierta habilidad de manera exhaustiva requiere la exploración de todas las secuencias de acciones, lo cual es una estrategia retrodroma. Es posible que la imprecisión de un experto al elaborar reactivos se deba a que está acostumbrado a resolver problemas siguiendo una estrategia prodroma, mientras que la determinación del contenido de un examen implica el empleo de una estrategia retrodroma.

Lewicki, Hiel y Bizet (1988) aportan evidencia de que el dominio de una hábilidad no garantiza que se es capaz de explicar en qué consiste esa habilidad. Dichos autores encontraron que la ejecución de sus sujetos en una tarea de predicción que implicaba el reconocimiento de un patrón de presentación de estímulos visuales, mejoraba notablemente con la práctica. A pesar de haber llegado a adquirir un conocimiento procedural preciso que les permitía predecir el patrón de estímulos que se les presentaba, los sujetos fueron incapaces de describir el procedimiento que empleaban.

Estas evidencias deben interpretarse como una prueba de que el empleo de expertos no es una condición suficiente para elaborar exámenes.

Las deficiencias señaladas parecen tener su origen en la ausencia de una ***base formal***: una manera de describir y analizar objetivamente el tema o la habilidad cuyo aprendizaje se desea evaluar, a fin de determinar el contenido de los reactivos.

Una buena base formal para describir y analizar un tema o habilidad puede radicar en la identificación de las ***propiedades estructurales***: las relaciones entre los elementos constituyentes de ese tema o esa habilidad (Solano-Flores, 1982, 1983). Al pretender evaluar habilidades académicas complejas, Cedeño y Ruiz-Primo (1982) identificaron un conjunto de elementos constituyentes que descompusieron en elementos más específicos, que a su vez descompusieron en otros más específicos, y así sucesivamente. Siguiendo esa lógica, Cedeño y Ruiz-Primo elaboraron una representación de las relaciones de inclusividad entre los elementos, que debiera tener un repertorio ideal, perfecto. Dicho repertorio ideal fue empleado como base de comparación para evaluar la ejecución de los alumnos: si un elemento no era identificado en la ejecución, se infería que tampoco estaban presentes los elementos que descendían de él (aquéllos a los que incluía). Por otra parte, si el elemento era identificado en la ejecución, se procedía a verificar la presencia de aquellos elementos que descendían de él. A pesar de que las pruebas evaluadas eran de ensayo, Cedeño y Ruiz-Primo alcanzaron una confiabilidad altísima en sus evaluaciones, lo que demuestra que la base formal que emplearon produjo una mayor objetividad en la evaluación.

Resulta claro que las propiedades estructurales son un factor digno de considerar al elaborar un examen: conociendo su organización, es posible determinar qué en concreto de un tema o una habilidad puede o debe ser evaluado. Ello constituye el primer acercamiento al contenido mismo de los reactivos.

El ***Diseño Lógico de Exámenes (DLE)*** es una técnica desarrollada con el propósito de derivar el contenido de los reactivos de un examen a partir de la identificación y del análisis de las relaciones entre los elementos componentes de una habilidad, y se ocupa de la evaluación del aprendizaje de procedimientos (Solano-Flores, en prensa). En virtud de sus propiedades formales, el DLE no solamente permite elaborar reactivos, sino que además ofrece interesantes posibilidades para la realización de tareas tales como la elaboración de reactivos de opción múltiple, la confección de exámenes paralelos y la construcción de bancos de reactivos.

Hasta ahora, el DLE carecía de la sustentación empírica que permitiera determinar sus alcances y limitaciones como recurso para identificar o predecir la dificultad de los reactivos. Ello, a pesar de que la técnica ofrece un conjunto de indicadores y medidas que pueden ser empleadas para estimar la complejidad de un reactivo.

Esta investigación tuvo el propósito de establecer la utilidad que algunos indicadores y medidas de la complejidad estructural de los reactivos, surgidos del DLE, tienen para predecir la dificultad de un reactivo. Asimismo, pretendió establecer la eficacia del DLE en la elaboración de exámenes paralelos. Antes de describir en qué consistió la investigación, es necesario revisar algunos conceptos e ideas fundamentales de la técnica.

F U N D A M E N T O S   Y   F A S E S   D E L   D I S E Ñ O   L O G I C O   D E   E X A M E N E S

El DLE consta de tres fases: ***1) formalización del procedimiento***, que permite describir las relaciones entre los elementos que constituyen un procedimiento; ***2) análisis de variables***, que permite identificar cuántos y qué tipos de reactivos se requieren para evaluar el aprendizaje de un procedimiento; y ***3) generación y análisis lógico de reactivos*** que, además de guiar la elaboración misma del reactivo, permite estimar y manipular la complejidad de los reactivos de un examen.

PRIMERA FASE: FORMALIZACION DEL PROCEDIMIENTO

Anderson (1976, 1985) distingue dos tipos de conocimiento: declarativo y procedural. El ***conocimiento declarativo*** se refiere a hechos y eventos que ocurren en el mundo; el ***conocimiento procedural*** se refiere a la manera de realizar actividades de tipo cognoscitivo. Tal diferenciación es útil para delimitar el ámbito de aplicación de la técnica que aquí será descrita: el DLE se ocupa de la evaluación del aprendizaje de procedimientos.

Por lo demás, los objetivos de ese autor y los de este trabajo son diferentes, aunque no incompatibles. Anderson (1983) se ha preocupado por el desarrollo de un modelo teórico de la organización del conocimiento y la estrategia de solución de problemas. En cambio, en este trabajo se presenta una forma de elaborar exámenes para evaluar el aprendizaje de procedimientos. El DLE se basa en la representación de los procedimientos, no en el proceso mental involucrado en su aplicación. Sin embargo, debe decirse que sus bases descriptivas y analíticas pueden ser empleadas para la elaboración de modelos de solución de problemas.

La principal característica de todo procedimiento es la existencia de una secuencia de pasos o decisiones. Los ***procedimientos*** han sido definidos como conjuntos de pasos o tareas que deben realizarse a fin de alcanzar una meta o llegar a un cierto producto (véase Trakhtenbrot, 1963; Landa, 1972, 1974). Roid y Haladyna (1982) definen a los procedimientos como secuencias de actividades físicas o mentales que se realizan para resolver problemas o alcanzar metas pertenecientes a una misma clase.

En adelante, cuando se hable de ***acción***, deberá entenderse cierta actividad física o mental concreta involucrada en un procedimiento. Básicamente, las acciones pueden ser operaciones o decisiones. La ***operaciones*** son actividades concretas que producen una transformación en el problema que se pretende resolver. Ejemplos de operación pueden ser: "suma los números de la columna", "desconecta el aparato", "subraya el título del libro", o "entrega tu solicitud en la oficina de admisiones".

Las ***decisiones*** son preguntas de cuya respuesta depende la secuencia de acciones que se sigue al aplicar el procedimiento. Las decisiones se han caracterizado como constructos que implican la solución de un conflicto para determinar los actos que a continuación deben ser ejecutados (Anderson, 1984). Ejemplos de decisión pueden ser: "¿queda algún número por calcular?", "¿el cociente es mayor que cero?", "¿el ciudadano es casado?", o "¿el paciente es alérgico a la penicilina?".

En el DLE, la formalización del procedimiento se basa en la representación apropiada de la secuencia en que deben realizarse las acciones. Además de describir las relaciones temporales entre los elementos, el objetivo de esta fase es identificar las propiedades estructurales del procedimiento.

En la Figura 1 aparece un sencillo procedimiento y su representación mediante lo que se denomina ***grafo*** (véase Harary, 1969). En un grafo, los elementos se representan mediante círculos o cuadros; las relaciones entre los elementos se representan mediante ***segmentos***. En el ejemplo

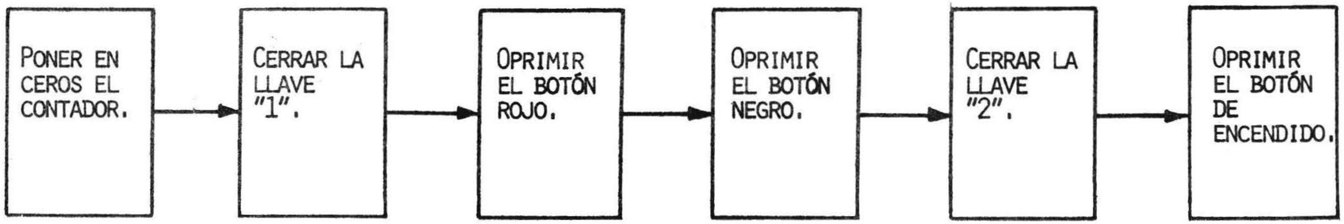


FIGURA 1. REPRESENTACIÓN DE UN PROCEDIMIENTO SENCILLO MEDIANTE UN GRAFO.

de la Figura 1, los segmentos son flechas que representan ***relaciones de secuencia*** o ***relaciones antes-despuès*** entre las acciones que prescribe el procedimiento.

Procedimientos más complejos, como el que aprece en la Figura 2, involucran la toma de decisiones, y pueden ser representados mediante diagramas de flujo, cuya elaboración ha sido ampliamente descrita por diversos autores (por ejemplo: Schriber, 1969; Farina 1970; Wheatley y Unwin, 1972; Stern, 1975; y Martin y McClure, 1985).

Gracias a los grafos y a los diagramas de flujo, es posible identificar tres propiedades estrucutrales que los procedimientos pueden tener. En la Tabla 1 se explican dichas propiedades estructurales. Un procedimiento es más complejo entre mas ***ramificaciones***, ***ciclos*** y ***circuitos*** contenga, ya que estas propiedades multiplican las posibles secuencias de acción al resolver un problema particular. Un ***procedimiento lineal***, como el que aparece en la Figura 1, no tiene ninguna de esas propiedades: consiste en una sola secuencia posible de acciones. En cambio, el procedimiento de la Figura 2 contiene varias secuencias posibles, cada una de las cuales es aplicable a una subclase especifica de problemas o metas: seguir una u otra secuencia depende de la naturaleza del problema especifico a resolver.

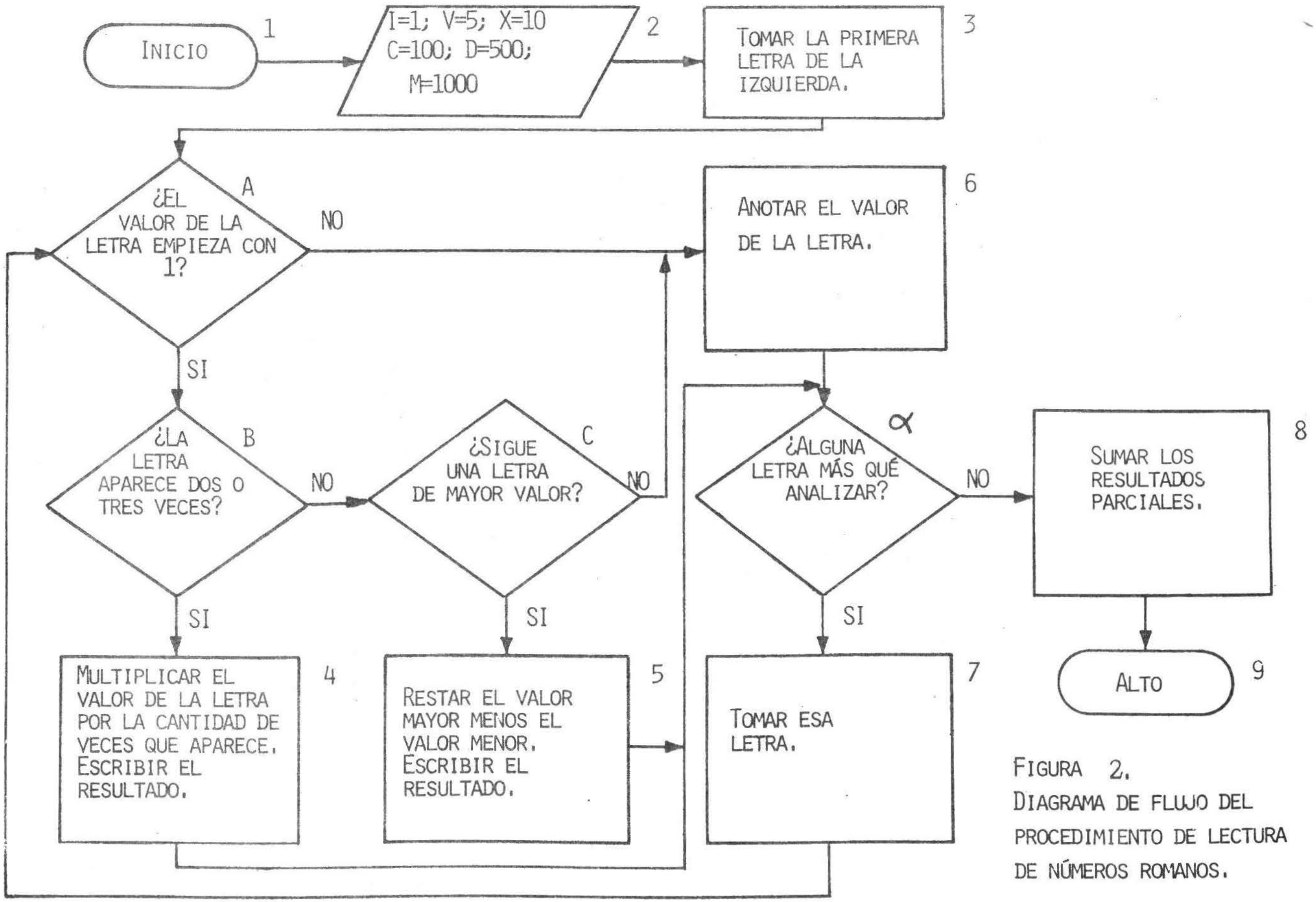
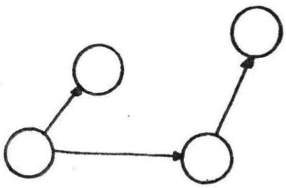
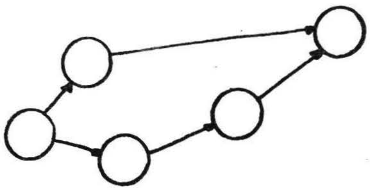
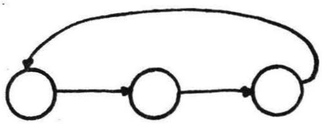


FIGURA 2.
DIAGRAMA DE FLUJO DEL PROCEDIMIENTO DE LECTURA DE NÚMEROS ROMANOS.

Tabla 1

Propiedades Estructurales que puede Haber en un Procedimiento

Ramificación		Dos o más secuencias posibles de acción.
Circuito		Dos o más secuencias de acciones convergen en un mismo elemento.
Ciclo		Se prescribe la realización de una acción ejecutada anteriormente.

La evaluación del aprendizaje de procedimientos sería muy sencilla si todos los procedimientos tuvieran una estructura lineal: puesto que la estructura lineal implica una sola secuencia posible de acciones, cualquier reactivo requeriría invariablemente esa secuencia para su correcta solución. Pero la presencia de ramificaciones, circuitos o ciclos en un procedimiento, implica la existencia de dos o más subclases de problema o meta.

SEGUNDA FASE: ANALISIS DE VARIABLES

En esta fase es posible determinar: 1) cuántas subclases de problema o meta puede manejar un procedimiento; 2) cuántas y cuáles secuencias posibles de acciones tiene el procedimiento; y 3) cuántos diferentes reactivos se requieren para evaluar el aprendizaje del procedimiento si se consideran todas esas posibles secuencias.

Obsérvese la Figura 3. En ella aparece un procedimiento que puede ser llamado ***procedimiento con estructura de árbol ramificado***: cada decisión que toma el usuario del procedimiento determina qué acciones han de llevarse a cabo. En ese procedimiento pueden distinguirse dos propiedades: 1) una vez separadas, dos ramas no vuelven a juntarse; 2) existen tantos elementos situados al final de una secuencia, como subclases de problemas a resolver o de metas a

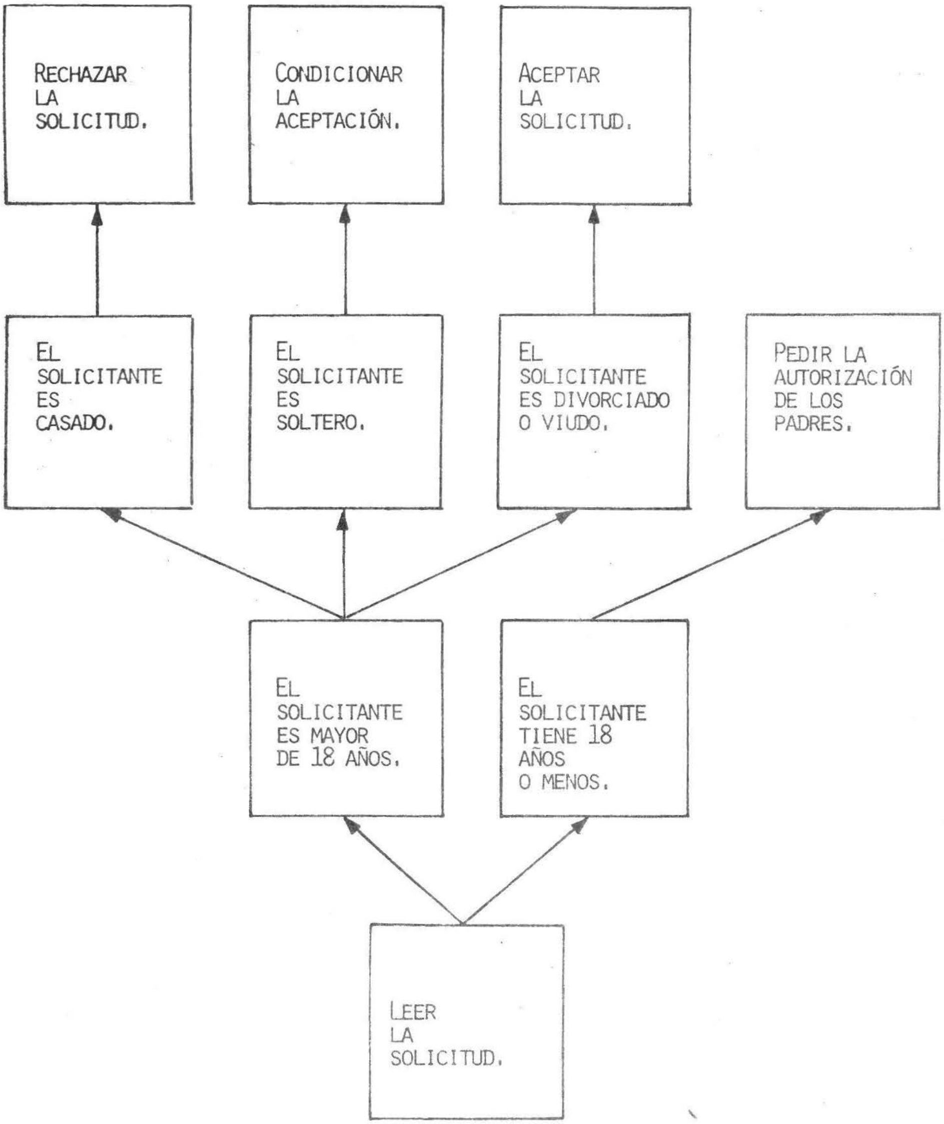


FIGURA 3. PROCEDIMIENTO CON ESTRUCTURA DE ÁRBOL RAMIFICADO.

alcanzar. Estos elementos reciben el nombre de *vèrtices cima* (Harary, 1969; Jackson y Gonzàlez, 1979; Salazar-Resines, 1979). La cantidad de vèrtices cima es igual a la cantidad de posibles trayectorias que contiene un procedimiento.

De acuerdo con la lògica expuesta, un reactivo que evalùa el aprendizaje de un procedimiento es representativo de una sola subclase de problema o meta.

Cuando un procedimiento tiene la apariencia de la Figura 3, es fàcil identificar las subclases de problema que maneja el procedimiento, asì como los diferentes tipos de reactivos que pueden ser elaborados. En casos asì, este es el momento en que conluye la fase de anàlisis de variables. Sin embargo, cuando un procedimiento no tiene la estructura de àrbol ramificado, esta fase se encarga adicionalmente de llegar a una representaciòn como la de la Figura 3.

Analìcese la Figura 2. La estructura del procedimiento ahì representado no es de àrbol ramificado: es difìcil identificar las subclases de problema que resuelve, por lo que no es fàcil determinar cuàntos y cuàles reactivos pueden incluirse en el examen que evalùe el aprendizaje de ese procedimiento.

El procedimiento de la Figura 2 contiene un circuito y un ciclo. El circuito se debe a la presencia de dos trayectorias que convergen en un mismo elemento (señalado con el número 6). Dichas trayectorias son: **A (opción NO)--->** 6; y **A** (opción SI) ---> **B (opción NO) ---> C (opción NO)** ---> 6.

El ciclo se debe al hecho de que, una vez cumplido el elemento 7, debe realizarse una acción (tomar la decisión A) previamente realizada. A fin de determinar cuántas diferentes subclases de problemas o metas maneja el procedimiento, éste debe ser "desmenuzado". Para lograrlo, es preciso distinguir entre dos tipos de decisiones:

1) ***Decisiones de reconocimiento*** (decisiones **A, B** y **C** en la Figura 2). Estas decisiones permiten al usuario del procedimiento identificar las características de un problema particular a resolver o de una meta particular a alcanzar. En adelante, las decisiones de reconocimiento se denotarán mediante letras latinas mayúsculas.

2) ***Decisiones de control*** (decisión α en la Figura 2). Estas decisiones tienen una opción (en el ejemplo, la opción **SI** de la decisión α) que prescribe la repetición de ciertas acciones una y otra vez hasta que se cumple cierta condición. En el ejemplo, las acciones se repiten una y otra vez hasta que sea posible contestar **NO** a la decisión α . En adelante,

las decisiones de control se denotarán mediante letras griegas mayúsculas.

Las decisiones de reconocimiento son cruciales para el DLE debido a que permiten identificar las características que debe tener cualquier reactivo. Como puede verse en la Figura 2, un reactivo que evalúa el aprendizaje del procedimiento de lectura de números romanos puede tener o no tener las siguientes características: la letra tiene un valor que debe ser escrito comenzando con un "1" (decisión A); la letra aparece dos o tres veces (decisión B); la letra es sucedida por otra letra de valor más alto (decisión C).

En el DLE, las decisiones de reconocimiento son consideradas como ***variables lógicas*** o ***binarias***, que pueden adoptar una de dos valores: positivo (SI) o negativo (NO). Una subclase de problema o meta es un conjunto específico de valores de variables binarias. Así, el análisis de variables sirve para identificar las relaciones entre las decisiones de reconocimiento y para identificar los posibles conjuntos de valores de variables binarias, a fin de deducir el contenido de los reactivos de un examen.

La fase de análisis de variables es prácticamente innecesaria en procedimientos como los de las Figuras 1 y 3. En casos como el de la Figura 2, la presencia de ciclos o circuitos hace necesaria la elaboración de un ***àrbol expansivo*** (vèase Harary, 1969). En la Figura 4 se presenta un àrbol expansivo del procedimiento para leer nùmeros romanos. Este àrbol se construyò a partir de la Figura 2. A fin de entender còmo se elabora un àrbol expansivo, compàrese a la Figura 4 con la Figura 2. Podràn notarse tres hechos:

1) En el àrbol expansivo no aparecen las leyendas de los elementos, lo que permite una representaciòn muy simple. A cada elemento se le denota mediante el nùmero o la letra con que se le denotò en el diagrama de flujo. Las letras se reservan para las decisiones; los nùmeros, para los elementos restantes.

2) Un mismo elemento puede aparecer dos o màs veces en el àrbol expansivo. De hecho, aparece igual cantidad de veces que las secuencias de acciones en que està involucrado.

3) En el àrbol expansivo, las decisiones no se representan como una totalidad. En vez de ello, cada opciòn se representa como si fuera un elemento diferente. Por ejemplo: **a** para la opciòn positiva de **A** y $\overline{\mathbf{a}}$ para la opciòn negativa de **A.** Se emplean letras minùsculas para denotar estas

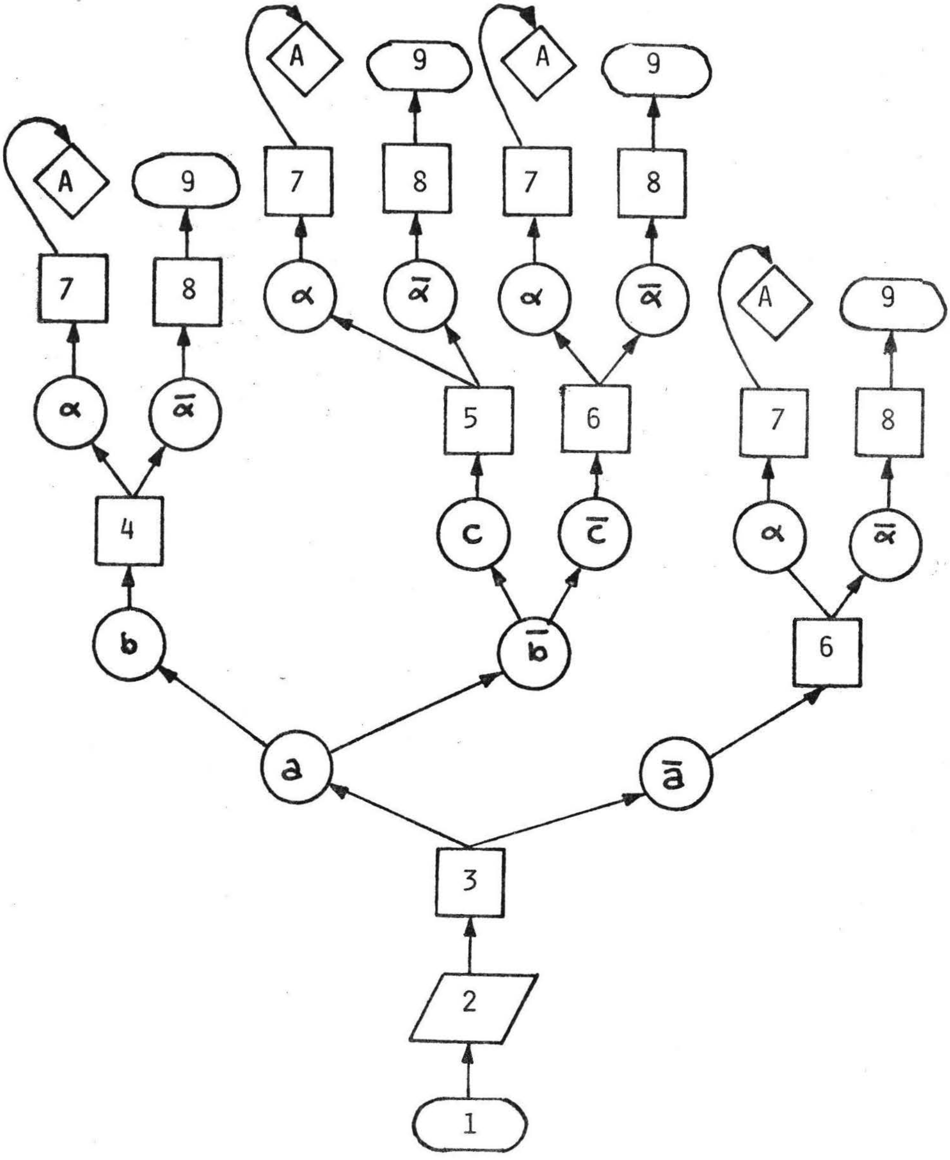


FIGURA 4. ÁRBOL EXPANSIVO DEL PROCEDIMIENTO PARA LA LECTURA DE NÚMEROS ROMANOS. EL ÁRBOL EXPANSIVO SE CONSTRUYÓ A PARTIR DEL DIAGRAMA DE FLUJO QUE APARECE EN LA FIGURA 2.

opciones.

4) En el árbol expansivo cada rama concluye en cualquiera de estos casos: cuando se llega al elemento **ALTO** (en el ejemplo, el elemento 9); cuando se llega a un ciclo. En el diagrama de flujo de la Figura 2, hay una flecha que parte del elemento 7 a la decisión **A**, planteada ya anteriormente. Sería reiterativo representar una y otra vez la secuencia de acciones que se repite después del elemento 7: ello conduciría a una repetición infinita en el árbol expansivo. Por tal razón, en la Figura 4 se ha trazado una flecha curva que van del elemento 7 a la decisión **A**.

El árbol expansivo permite enumerar y describir todas las posibles subclases de problemas o metas que contiene un procedimiento en el que hay circuitos o ciclos. En el árbol expansivo, cada clase corresponde a una ramificación que concluye con el elemento terminal **ALTO**. Si se sigue lo prescrito por el diagrama de flujo de la Figura 2 y se identifica la trayectoria correspondiente en el árbol expansivo de la Figura 4, podrá comprobarse que la lectura de los números romanos **XL** (40) y **IX** (9), requiere seguir exactamente la misma secuencia de acciones:

$$1 \text{--->} 2 \text{--->} 3 \text{--->} a \text{--->} \bar{b} \text{--->} c \text{--->} 5 \text{--->} \bar{\alpha} \text{--->} 8 \text{--->} 9.$$

El conjunto de valores de variables: **a**, $\overline{\mathbf{b}}$ y **c**, define la naturaleza de los reactivos: "¿Qué número es **XL**?" y "¿Qué número es IX?". Puede, en consecuencia, enunciarse el principio general de que dos reactivos cualesquiera cuya solución involucre al mismo conjunto de valores de variables lógicas al aplicar el procedimiento, pertenecerán a la misma subclase de problemas o metas. Es el caso en que puede decirse que dos reactivos formalmente equivalentes. Tal principio es la base para elaborar exámenes paralelos mediante el DLE. Mediante el grafo (cuando el procedimiento tiene una estructura lineal o de árbol ramificado) o el árbol de expansión (cuando el procedimiento contiene circuitos o ciclos), puede decidirse cuál conjunto de valores de variables deben ser considerados al elaborar el examen, y se pueden crear o seleccionar reactivos que correspondan a esos conjuntos de valores de variables. En consecuencia, la elaboración de reactivos puede ser vista como la manipulación de conjuntos de valores de variables lógicas.

TERCERA FASE: GENERACION Y ANALISIS LOGICO DE REACTIVOS

Un reactivo puede ser descrito sintéticamente mediante los valores de las variables lógicas que toman parte en su solución. Se han adoptado dos convenciones para su descripción:

1) Usar el número **1** para denotar el valor positivo de **la** variable (opción **SI** de la decisión); usar **-1** para **denotar el valor negativo** de la variable (opción **NO** **de la** decisión); y usar **0** para denotar la ausencia de una variable (lo que significa que en la trayectoria seguida al resolver el reactivo no aparece la decisión correspondiente; **0** significa: "ni **1** ni **-1**").

2) Usar el siguiente modo de representación de los valores de las variables lógicas:

ARR = (**A** **B** **C** ... K).

Esta forma de representación es llamada ***arreglo*** o ***arreglo simple*** (***ARR***, de ahora en adelante), y se define como la especificación de valores de un conjunto de valores de variables (véase Wirth, 1976). **A**, **B**, **C**, ..., K, son variables lógicas o binarias.

Según lo anterior, los reactivos formalmente equivalentes: "¿Qué número es XL?" y "¿Qué número es **IX**?", son descritos por el mismo arreglo:

$$\begin{array}{cccc} & A & B & C \\ ARR = (& 1 & -1 & 1 \ ) . \end{array}$$

Puede constatarse en la Figura 4 que sólo hay cuatro trayectorias que culminan en el elemento terminal **ALTO:** **(1 1 0), (1 -1 1), (1 -1 -1)** y **(-1 0 0).** No es necesario escribir las letras **A**, **B**, **C**, etcétera, cuando se conoce el lugar que cada variable tiene asignada en el arreglo.

La cantidad de arreglos diferentes es igual a la cantidad de subclases de problemas o metas que un procedimiento maneja. En el DLE, los reactivos se generan buscando o seleccionando aquellos problemas que satisfacen a las descripciones de los arreglos. El análisis lógico de reactivos (no confundir con el análisis de reactivos) consiste en: 1) resolver los reactivos de acuerdo con el procedimiento según se le formalizó en la primera fase; y 2) identificar la trayectoria seguida al resolverlo y, por tanto, el arreglo que describe al reactivo.

El conocimiento de los diferentes arreglos que maneja un procedimiento permite agrupar y diferenciar a los reactivos de acuerdo con su composición estructural. Gracias a ello, es posible determinar: 1) qué, exactamente, evalúa cada reactivo; 2) cuales son los arreglos que deben ser considerados dependiendo de los propósitos del evaluador; 3) cuáles son los reactivos que deben incluirse en el examen; y 4) qué reactivos son formalmente equivalentes.

El DLE no proporciona solamente criterios cualitativos para analizar a los reactivos. Existen otras nociones que pueden ayudar a estimar la complejidad de un reactivo, y a agrupar y diferenciar a los reactivos de acuerdo con criterios cuantitativos. Una de esas nociones es la de ***dimensión*** (***DIM***, en adelante), que se define como la cantidad de variables presentes en un ARR.

Hasta ahora, al exponer el ejemplo del procedimiento de lectura de números romanos, se han omitido los casos en que, debido a la decisión de control α (véase la Figura 2), el procedimiento se aplica varias veces al resolver un mismo reactivo. La lectura de números romanos es un procedimiento que pertenece a la categoría de ***procedimientos cerrados***, en los cuales, para resolver ciertos problemas, es necesario volver a aplicar el procedimiento, aunque no necesariamente la misma secuencia. Leer un número romano como XLV implica seguir la siguiente trayectoria:

$$1 \to 2 \to 3 \to a \to \bar{b} \to c \to 5 \to \alpha \to 7 \to \bar{a} \to 6 \to \bar{\alpha} \to 8 \to 9 \ .$$

Puede comprobarse que la decisión de control α aparece dos veces en la secuencia (primero la opción α , después la opción $\bar{\alpha}$). En consecuencia, al leer el número romano XLV se han dado dos "vueltas" en el procedimiento. A tales vueltas se les llama ***iteraciones***. En la primera iteración, se resuelve la parte **XL** (40) del reactivo, mediante el **ARR** = (1 -1 1); en la segunda iteración se resuelve el resto del reactivo, **V** (5), mediante el **ARR= (-1 0 0)**. Una vez ejecutadas tales iteraciones, se sigue la opción **NO** de α, en virtud de que no se encuentran más letras què leer. El siguiente paso es sumar los dos resultados parciales (40 y 5), para obtener el resultado total.

El ***arreglo múltiple*** de un reactivo (**ARR MUL**, en adelante) es la descripción que enumera los arreglos simples en el mismo orden en que se presentan al resolver el reactivo. El reactivo: "¿Qué número es **XLV?**", tiene el siguiente ARR MUL:

ARR MUL = (1 -1 1) (-1 0 0) .

La Tabla 2 presenta diversos arreglos múltiples. Puede apreciarse que existen reactivos què, a pesar de tener apariencia muy diferente, tienen el mismo arreglo. Tal es el caso de los reactivos **DI** (501) y **VI** (6). Ello significa que requieren exactamente la misma secuencia de acciones del procedimiento para llegar a una solución correcta. En el DLE, los reactivos formalmente equivalentes son aquèllos que tienen un mismo arreglo.

Tabla 2

Diversos Arreglos y Reactivos

ARR o ARR MUL	REACTIVOS
(-1 0 0)	V;L;D
(1 -1 -1)	X;C;M
(-1 0 0)(-1 0 0)	LV;DV;DL
(-1 0 0)(1 -1 -1)	DI;DX;DC;LI; LX;VI
(1 -1 -1)(1 -1 -1)(1 -1 -1)	CXI;MCX;MCI; MXI
(1 -1 1)(1 -1 1)(-1 0 0)	CDXLV;CMCXCV
(-1 0 0)(1 1 0)(1 -1 -1)	LXXI;LXXXI;DXXI; DXXXI;DCCI;DCCCI
(1 -1 -1)(1 -1 -1)(1 -1 -1)(1 -1 -1)	MCXI
(1 -1 -1)(1 -1 -1)(1 -1 -1)(-1 0 0)	MCXV
(-1 0 0)(1 -1 -1)(1 -1 -1)(1 -1 -1)	DCXIV
(1 -1 -1)(-1 0 0)(1 -1 1)(1 1 0)	MDXCII

El ARR MUL de un reactivo no sólo permite identificar si dos reactivos cualesquiera son formalmente equivalentes. También permite identificar las siguientes variables:

1) ***Iteratividad*** (*ITE*, en adelante), definida como el número de iteraciones involucradas en la solución del reactivo.

2) ***Grado múltiple*** (*GRA MUL*, en adelante), definida como el número de arreglos simples distintos involucrados en la solución del reactivo.

3) ***Decisiones acumuladas*** (*DEC ACU*, en adelante), definida como la cantidad de veces que se toma una decisión cualquiera al resolver un reactivo. Puede calcularse como la cantidad de **unos** positivos y negativos que aparecen en el ARR MUL del reactivo.

4) ***Pasos acumulados*** (*PAS ACU*), en adelante), definida como la cantidad de veces que "se pasa" por algún elemento , al resolver el reactivo.

En realidad, PAS ACU y DEC ACU están íntimamente asociadas a ITE. En la Figura 5 aparece la relación entre ITE y PAS ACU, y en la Figura 6 aparece la relación entre ITE y DEC ACU. Ambas relaciones se obtuvieron a partir de un

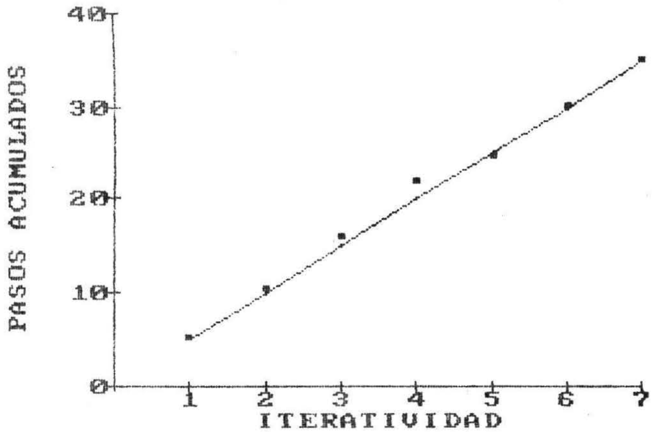


FIGURA 5. RELACIÓN ENTRE LA ITERATIVIDAD Y LA CANTIDAD DE PASOS ACUMULADOS.

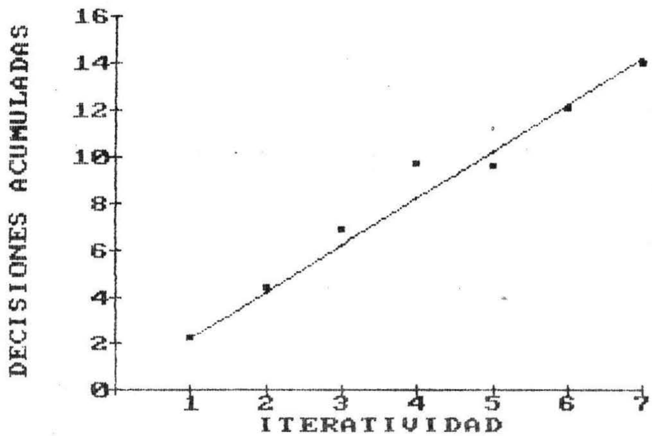


FIGURA 6. RELACIÓN ENTRE LA ITERATIVIDAD Y LA CANTIDAD DE DECISIONES ACUMULADAS.

sondeo previo a la elaboración de este proyecto, en el que se exploraron las distintas variables que podían ser manipuladas al elaborar un reactivo. En tal sondeo se obtuvo una muestra de 53 reactivos de diferente ITE y GRA MUL, a los que se aplicó el análisis lógico de reactivos. Gracias al análisis lógico de reactivos pudo estimarse el valor de PAS ACU y DEC ACU de cada reactivo.

Se aprecia que las relaciones son lineales y directas, de lo que se concluye que al manipular a ITE se manipula a PAS ACU y a DEC ACU.

U N A   I N V E S T I G A C I O N   P A R A   E S T I M A R   L A   C A P A C I D A D   P R E D I C T I V A   D E L   D I S E N O   L O G I C O   D E   E X A M E N E S

OBJETIVO DE LA INVESTIGACION

De lo expuesto anteriormente es fácil concluir que el DLE tiene ciertas propiedades formales y analíticas que hacen más objetivo y sistemático el trabajo de elaboración de reactivos. Sin embargo, debido a que la técnica carecía aún del sustento empírico suficiente para recomendar su aplicación en situaciones prácticas, se pretendió determinar la capacidad predictiva de las variables surgidas del DLE.

Ello implicaba establecer la relación entre las variables ITE y GRA MUL, y la variable ***dificultad de un reactivo*** (***DIF***, en adelante), definida como el porcentaje de estudiantes que resuelven incorrectamente el reactivo. Así, la presente investigación tuvo como objetivo evaluar la capacidad que tienen las variables ITE y GRA MUL para predecir DIF. Adicionalmente, se pretendió comprobar si los reactivos identificados como formalmente equivalentes, según los principios del DLE (aquellos reactivos cuyo arreglo múltiple es el mismo) producen ejecuciones equivalentes entre los estudiantes.

De esta manera, dos hipótesis fueron sometidas a prueba:

H1: La complejidad de un reactivo, estimada según sus valores de iteratividad y grado múltiple, produce una mayor dificultad empírica del reactivo: existe una relación directa entre las variables iteratividad y grado múltiple, y la dificultad de un reactivo.

H2: Los reactivos formalmente equivalentes son resueltos por los estudiantes de manera equivalente. En otras palabras, dos reactivos con un mismo arreglo, tienen un alto grado de asociación: son resueltos ambos correctamente o ambos incorrectamente por una proporción significativa de estudiantes.

METODO

Variables. Se manipularon las variables independientes ITE y GRA MUL y la variable dependiente DIF, definidas anteriormente. Para estimar el grado de asociación de los reactivos cuyo arreglo fuera el mismo, se empleó el coeficiente ***phi***, que permite estimar el grado de asociación de dos variables dicotómicas (véase Magnusson, 1976; Levin, 1979) y que en esta investigación permitió estimar qué tan significativa era la proporción de alumnos que resolvía dos reactivos ambos correctamente o ambos incorrectamente.

Sujetos. Se empleó una muestra de 211 sujetos de ambos sexos que en el momento de correr la investigación cursaban el sexto año de la escuela primaria, y cuyas edades oscilaban entre los 11 y 13 años. Los sujetos empleados pertenecían a cuatro escuelas particulares mixtas de la Colonia del Valle, en la Ciudad de México.

Material y Procedimiento. Se empleó un examen consistente en un conjunto de cifras escritas en números romanos, que los sujetos debían escribir en la notación arábiga convencional.

Se eligió al procedimiento de lectura de números romanos por la razón de que su enseñanza forma parte de los contenidos oficiales del nivel de escuela primaria, de modo que no era necesario un entrenamiento previo de los sujetos. Adicionalmente, un examen de esta naturaleza ofrecía la ventaja de que su calificación es objetiva.

Para elaborar el examen, se tomaron los números comprendidos entre el 1 y el 2000, se les escribió en números romanos, y se les resolvió de acuerdo con el algoritmo que aparece en la Figura 2. Se efectuó el análisis lógico de cada reactivo, a fin de conocer su ARR MUL y determinar sus valores de ITE y GRA MUL.

Se obtuvo una muestra intencional de reactivos en la que estuvieran representadas todas las posibles combinaciones resultantes de valores de ITE y GRA MUL. Al hacerlo, se procuró tener representados a todos los posibles arreglos simples. Para facilitar la descripción de los arreglos múltiples, se recurrirá a la siguiente convención:

a	representa al arreglo sencillo:	(1 -1 -1)
b	" " " "	(1 1 0)
e	" " " "	(1 -1 1)
f	" " " "	(-1 0 0)

De esta manera, por ejemplo, el ARR MUL:

(1 -1 1) (1 -1 -1) (-1 0 0)

puede ser descrito por la serie de letras:

e a f .

En la Tabla 3 aparecen los reactivos seleccionados. Se aprecia que del total de 2000 reactivos no hubo ninguno con ITE mayor que 7. Tampoco hubo algún reactivo con GRA MUL mayor que 4, debido a que el procedimiento solamente tiene cuatro arreglos simples. Los cuadros vacíos corresponden a combinaciones de valores de ITE y GRA MUL que no existen.

Al obtener la muestra de reactivos, se procuró que en éstos estuvieran representados equitativamente los cuatro diferentes arreglos simples. La mayoría de las celdas de la Tabla 2 están ocupadas por cuatro reactivos: ello obedece a que se procuró balancear el orden en que aparecen los arreglos simples. Por ejemplo, en la celda correspondiente a ITE = 4 y GRA MUL = 3, el ARR MUL de un reactivo empieza con el arreglo simple **a**, el ARR MUL de otro reactivo empieza con el arreglo simple **b**, el ARR MUL de otro reactivo empieza con el arreglo simple **e**, y el ARR MUL del reactivo restante empieza con el arreglo sencillo **f**.

La razón de que existan celdas ocupadas con menos de cuatro reactivos, es que entre los números **1** y **2000**, hubo menos de cuatro reactivos que tuvieran esa combinación de valores de ITE y GRA MUL.

Tabla 3

REACTIVOS DEL EXAMEN

ITERATIVIDAD	Grado Múltiple 1	2
1	a C (33) b MM(11) e XC(62) f V(15)	
2	aa MI(9) bb CCCXXX(42) ee CMIX(28) ff DV(23)	ae CXL(30) be CCIV(10) ef CML(24) fb VII(1)
3	aaa MXI(25) bbb CCXXXII(13) eee CMXLIV(54) fff DLV(12)	aea MCDX(56) bfb CCCLII(65) ebe CMXXXIV(18) fee DXLIV(37)
4	aaaa MCXI(53)	aaee MCXCIX(8) bbfb CCXXXVII(7) fafa DCLX(55)
5		afafa MLXVI(6) bfbfb CCCLXXXVIII(50) fbbfb DCCCXXVIII(49)
6		afafaa MDCLXI(21) fbfbfb DCCLXXXVII(2)
7		afafafa MDCLXVI(4)

Las letras minúsculas describen el arreglo múltiple. Los números entre paréntesis indican la ubicación de los reactivos en el examen.

(continúa)

Tabla 3 (continuación)

REACTIVOS DEL EXAMEN

Grado Múltiple

ITERATIVIDAD	3	4
1		
2		
3	aef MCMV(60) bfe CCCLIX(34) eba CDXXXI(19) fab LXII(26)	
4	afee MDXCIV(32) befb CCXCVIII(46) eefa CMXCVI(14) fbfa DCCLX(59)	abef MCCXCV(57) befa CCXCVI(16) eafb CDXVII(40) faeb DCXCII(31)
5	abfbb MCCLXXXIII(68) bfbfa CCLXXXVI(35) efafa CMLXVI(41) ffbfa DLXXXVI(38)	afbee MDCCXCIX(51) efbfa CDLXXXVI(58) fbfae DCCLXIX(66)
6	aafbfa MCLXXXVI(39) fbfafa DCCCLXVI(44)	afbfae MDCCCLXIV(61)
7	afbfbfa MDCCCLXXVI(69)	

Las letras minúsculas describen el arreglo múltiple. Los números entre paréntesis indican la ubicación de los reactivos en el examen.

De los reactivos seleccionados para cada combinación de valores de ITE y GRA MUL, se seleccionó al azar uno para que en el examen tuviera un reactivo equivalente; es decir, otro reactivo con el mismo arreglo múltiple.

En la Tabla 4 aparecen los reactivos formalmente equivalentes del examen que se empleó.

Según se puede comprobar en las la Tablas 3 y 4, el examen empleado estuvo constituido por 69 reactivos diferentes. El examen, tal como se le empleó, se reproduce en la Figura 7.

A fin de evitar que la práctica o el cansancio ejercieran algún efecto sobre la ejecución de los sujetos, se asignó aleatoriamente el orden de presentación de cada reactivo en el examen. Se observó que los alumnos no seguían un orden uniforme en la contestación de los reactivos: algunos los resolvían procediendo de arriba hacia abajo, otros de izquierda a derecha, otros en un orden inespecifico; ello constituyó de algún modo un control por aleatorización del orden de presentación de los reactivos. Adicionalmente, se sumaron las frecuencias de soluciones incorrectas según los reactivos estuvieran localizados en la parte superior, media o inferior de la hoja de examen, y a continuación se efectuó un análisis de varianza: se encontró que la cantidad de soluciones incorrecetas no varió

Tabla 4

REACTIVOS EQUIVALENTES EN EL EXAMEN

Grado Múltiple

ITERATIVIDAD		1	2
	1	b MM(11) b III(20)	
	2	ff DV(23) ff DL(27)	fb VII(1) fb DCC(29)
	3	bbb CCXXXII(13) bbb CCCXXII(5)	bfb CCCLII(65) bfb XXXVIII(22)
	4		aaee MCXCIX(8) aaee MCXLIX(17)
	5		bfbfb CCCLXXXVIII(50) bfbfb CCCLXXVII(43)
	6		fbfbfb DCCLXXXVII(2) fbfbfb DCCLXXVII(36)
	7		

(continúa)

Las letras minúsculas describen el arreglo múltiple. Los números entre paréntesis indican la ubicación de los reactivos en el examen.

En cada celda aparece, en primer lugar el reactivo incluido en la Tabla 3. El reactivo que aparece en segundo lugar es el reactivo "extra" que se creó para que fuera equivalente del primero.

Tabla 4 (continuación)

REACTIVOS EQUIVALENTES EN EL EXAMEN

ITERATIVIDAD	Grado Múltiple 3		Grado Múltiple 4	
1				
2				
3	fab fab	LXII(26) DXII(48)		
4	befb befb	CCXCVIII(46) CCXLVIII(64)	eafb eafb	CDXVII(40) CMXVIII(47)
5	efafa efafa	CMLXVI(41) CDLXVI(45)	efbfa efbfa	CDLXXXVI(58) CDLXXVI(63)
6	aafbfa aafbfa	MCLXXXVI(39) MCLXXVI(52)	afbfae afbfae	MDCCCLXIV(61) MDCCCLXIX(3)
7	afbfbfa afbfbfa	MDCCCLXXVI(69) MDCCLXXVI(67)		

Las letras minúsculas describen el arreglo múltiple. Los números entre paréntesis indican la ubicación de los reactivos en el examen.

En cada celda aparece, en primer lugar, el reactivo incluido en la Tabla 3. El reactivo que aparece en segundo lugar es el reactivo "extra" que se creó para que fuera equivalente del primero.

INSTRUCCIONES

A continuación hay una lista de números romanos. Al lado de cada uno, anota en números arábigos, la cantidad que pienses que está representada.

VII________	DCCLXXXVII__________	MDCCCLXIX__________
MDCLXVI__________	CCCXXII__________	MLXVI__________
CCXXXVII__________	MCXCIX__________	MI__________
CCIV__________	MM__________	DLV__________
CCXXXII__________	CMXCVI__________	V__________
CCXCVI__________	MCXLIX__________	CMXXXIV__________
CDXXXI__________	III__________	MDCLXI__________
XXXVIII__________	DV__________	CML__________
MXI__________	LXII__________	DL__________
CMIX__________	DCC__________	CXL__________
DCXCII__________	MDXCIV__________	C__________
CCCLIX__________	CCLXXXVI__________	DCCLXXVII__________
DXLIV__________	DLXXXVI__________	MCLXXXVI__________
CDXVII__________	CMLXVI__________	CCCXXX__________
CCCLXXVII__________	DCCCLXVI__________	CDLXVI__________
CCXCVIII__________	CMXVIII__________	DXII__________
DCCCXXVIII__________	CCCLXXXVIII__________	MDCCXCIX__________
MCLXXVI__________	MCXI__________	CMXLIV__________
DCLX__________	MCDX__________	MCCXCV__________
CDLXXXVI__________	DCCLX__________	MCMV__________
MDCCCLXIV__________	XC__________	CDLXXVI__________
CCXLVIII__________	CCCLII__________	DCCLXIX__________
MDCCLXXVI__________	MCCLXXXIII__________	MDCCCLXXVI__________

FIGURA 7. EXAMEN EMPLEADO EN LA INVESTIGACIÓN.

significativamente ($p < .05$), de modo que pudo tenerse la certeza de que la dificultad de los reactivos no estuvo influida por su ubicación en la hoja del examen.

El examen se aplicó de manera colectiva, simultáneamente en todos los grupos de cada escuela, de modo que no hubo posibilidad de que algún alumno supiera de antemano acerca del examen o de su contenido. No hubo restricciones en el tiempo de aplicación. Simplemente se pidió a los alumnos no dejar ningún reactivo sin resolver, contestar el examen lo mejor que pudieran, y alzar la mano cuando terminaran, para que se les recogiera.

Para poner a prueba la hipótesis de que la mayor complejidad de un reactivo produce una mayor dificultad, se determinó la relación entre ITE y DIF, entre GRA MUL y DIF, y los efectos combinados de ITE y GRA MUL sobre DIF.

La hipótesis fue puesta a prueba según el siguiente modelo lineal:

$$DIF = A + B1\ (ITE) + B2\ (GRA\ MUL) \qquad (1),$$

en donde ***B1*** y ***B2*** son los coeficientes que se pretendía determinar.

Para el análisis de datos se procedió a calificar el examen de cada sujeto. Se contó la cantidad de alumnos que resolvió incorrectamente cada uno de los reactivos que aparecen en la Tabla 3. Se calculó a DIF como el porcentaje de alumnos que resolvió incorrectamente cada reactivo, y se relacionó a estos datos con los valores de ITE y de GRA MUL, mediante el análisis de regresión múltiple, que es el más apropiado para evaluar la capacidad predictiva del modelo lineal, y con el que es posible evaluar el efecto individual y combinado de las variables independientes sobre la variable dependiente.

Para poner a prueba la hipótesis de que los reactivos formalmente equivalentes son resueltos ambos correctamente o ambos incorrectamente por una gran proporción de alumnos, solamente se tomó en cuenta a los reactivos de la Tabla 4. Se calculó el coeficiente ***phi*** para cada par de reactivos formalmente equivalentes. Ya que el examen contenía 16 pares de reactivos con esas características, se obtuvo la misma cantidad de coeficientes phi. Se puso a prueba la significatividad de cada coeficiente phi mediante su contrastación con la distribución de valores de ***chi*** cuadrada.

R E S U L T A D O S

Efectos de la iteratividad y del grado múltiple sobre la dificultad de un reactivo

El análisis de regresión múltiple se realizó mediante la modalidad de inclusión jerárquica. De acuerdo con esta modalidad, se calculó primero el efecto de ITE y posteriormente el efecto combinado de ITE y GRA MUL. Los resultados obtenidos se sintetizan en la Tabla 5.

Se puede apreciar un valor de R cuadrada = 0.34013 cuando en la ecuación sólo se considera a ITE. Cuando la ecuación considera a ITE y GRA MUL, el valor de R cuadrada = 0.34027. La diferencia entre ambos variables es despreciable (14 diezmilésimas), lo que significa que el efecto combinado de ITE y GRA MUL no difiere significativamente del efecto aislado de ITE.

En la Tabla 6 puede constatarse que los coeficientes calculados para ITE (B1 = 2.4962; Beta 1 = 0.57516), muestran una magnitud considerable en comparación con los de GRA (B2 = 0.09487; Beta 2 = 0.01439).

También en la Tabla 6 puede notarse que el error estándar de ITE es menor que el de GRA MUL, por lo que las predicciones de los valores de DIF son más precisas cuando se considera a la variable ITE. Adicionalmente, el efecto de

Tabla 5

EFECTOS DE LAS VARIABLES CONSIDERADAS SOBRE LA DIFICULTAD DE LOS REACTIVOS

Variable	R múltiple	R cuadrada	Cambio en R cuadrada	R simple
ITE	0.58321	0.34013	0.34013	0.58321
GRA MUL	0.58333	0.34027	0.00014	0.33594

Tabla 6

COEFICIENTES B Y BETA DE LAS VARIABLES CONSIDERADAS EN LA ECUACION

Variable	B	Beta	Error estándar de B	Valores F
ITE	2.496282	0.57516	0.60128	17.236 *
GRA MUL	0.094879	0.01439	0.91322	0.011 +
(Constante)	2.903982			

* $p < .005$

+ no significativa

GRA MUL no es significativo ($F = 0.011$): las variaciones de DIF no son atribuibles a GRA MUL. Por contraste, ITE produjo un valor de $F = 17.236$ ($p < .005$).

Como puede verse en la Tabla 7, DIF e ITE tienen una correlación mayor ($r = 0.58321$) que DIF y GRA ($r = 0.33594$), lo que también demuestra que la influencia de ITE es más determinante de las variaciones de DIF.

Ya que ITE y GRA muestran cierta correlación (0.55906), se procedió a determinar el coeficiente de correlación semiparcial de ITE y GRA. Tal coeficiente permite determinar la influencia de una variable independiente, después de eliminar los efectos de la otra variable independiente y los efectos combinados de ambas variables independientes. Los coeficientes se obtuvieron de acuerdo con Cohen y Cohen (1975), mediante las siguientes fórmulas:

$$sr_{ITE} = \frac{r_{DIF*ITE} - (r_{DIF*GRA} \; r_{ITE*GRA})}{\sqrt{1 - (r_{ITE*GRA})^{2}}} \qquad (2)$$

$$sr_{GRA} = \frac{r_{DIF*GRA} - (r_{DIF*ITE} \; r_{ITE*GRA})}{\sqrt{1 - (r_{ITE*GRA})^{2}}} \qquad (3),$$

en donde sr_{ITE} es el coeficiente de correlación semiparcial de ITE; sr_{GRA} es el coeficiente de correlación de GRA; $r_{DIF*ITE}$ es el coeficiente de correlación entre DIF e ITE; $r_{DIF*GRA}$ es el coeficiente de correlación entre DIF y GRA; y $r_{ITE*GRA}$ es el coeficiente de correlación entre ITE y GRA.

Los coeficientes obtenidos mediante las ecuaciones (2) y (3) fueron: $sr_{ITE} = 0.4769$ y $sr_{GRA} = 0.0105$. De esta manera, puede decirse que una gran proporción de la variación en la dificultad de los reactivos es atribuible a los efectos aislados de ITE, mientras que los efectos aislados de GRA son mínimos.

En virtud de la casi nula influencia de GRA MUL, los valores del coeficiente B1 de ITE y de la constante de la ecuación son ligeramente mayores cuando se aislan los efectos de ITE. Ello puede comprobarse en la Tabla 8, en donde además se aprecia que el error estándar de ITE es menor (0.49368) y el valor de F es mayor ($F = 26.288$; $p < .005$), cuando se consideran los efectos aislados de esa variable.

Por lo anterior, puede concluirse que de las dos variables independientes probadas, ITE es la única que puede ser empleada como predictor de DIF. Según los datos que proporciona la Tabla 8, la función encontrada es:

en donde sr_{ITE} es el coeficiente de correlación semiparcial de ITE; sr_{GRA} es el coeficiente de correlación de GRA; $r_{DIF*ITE}$ es el coeficiente de correlación entre DIF e ITE; $r_{DIF*GRA}$ es el coeficiente de correlación entre DIF y GRA; y $r_{ITE*GRA}$ es el coeficiente de correlación entre ITE y GRA.

Los coeficientes obtenidos mediante las ecuaciones (2) y (3) fueron: $sr_{ITE} = 0.4769$ y $sr_{GRA} = 0.0105$. De esta manera, puede decirse que una gran proporción de la variación en la dificultad de los reactivos es atribuible a los efectos aislados de ITE, mientras que los efectos aislados de GRA son mínimos.

En virtud de la casi nula influencia de GRA MUL, los valores del coeficiente B1 de ITE y de la constante de la ecuación son ligeramente mayores cuando se aislan los efectos de ITE. Ello puede comprobarse en la Tabla 8, en donde además se aprecia que el error estándar de ITE es menor (0.49368) y el valor de F es mayor ($F = 26.288$; $p < .005$), cuando se consideran los efectos aislados de esa variable.

Por lo anterior, puede concluirse que de las dos variables independientes probadas, ITE es la única que puede ser empleada como predictor de DIF. Según los datos que proporciona la Tabla 8, la función encontrada es:

Tabla 7

COEFICIENTES DE CORRELACION ENTRE PARES DE VARIABLES

Variable	ITE	GRA	DIF
ITE	1.00000		
GRA	0.55906	1.00000	
DIF	0.58321	0.33594	1.00000

Tabla 8

COEFICIENTES B Y BETA PRODUCIDOS
POR LA VARIABLE ITERATIVIDAD

Variable	B	Beta	Error estándar de B	Valor F
ITE	2.531207	0.58321	0.49368	26.288 *
(Constante)	2.995492			

* $p < .005$

$$DIF = 2.995 + 2.531\ (ITE) \qquad (4)$$

En la Figura 8 aparecen la curva de esa función lineal y el diagrama de dispersión de los datos.

Ejecución de los alumnos ante reactivos formalmente equivalentes

Mediante el análisis CROSSTABS se elaboró, para cada par de reactivos *i* y *j* con un mismo arreglo múltiple, una tabla de doble entrada como la que aparece en la Figura 9. En dicha tabla, la celda **A** especifica cuántos alumnos resolvieron equivocadamente *i* y *j*; la celda **B** especifica cuántos resolvieron mal *i* y bien *j*; la celda **C** especifica cuántos resolvieron bien *i* y mal *j*; y la celda **D** especifica cuántos alumnos resolvieron bien *i* y *j*.

A partir de tablas de ese tipo, es posible calcular el valor de phi para cada para de reactivos con un mismo arreglo múltiple, de acuerdo con la siguiente fórmula:

$$phi = \sqrt{chi\ cuadrada\ /\ N} \qquad (5).$$

Por su parte, el valor de chi cuadrada se obtiene mediante la fórmula:

$$\textit{chi cuadrada} = \frac{N\,[(AD) - (BC)]^2}{(A+C)\;(B+D)\;(A+B)\;(C+D)} \qquad (6),$$

en donde N = 211, tamaño de la muestra de alumnos que se empleó.

En la Tabla 9 se enlistan los 16 pares de reactivos formalmente equivalentes incluidos en el examen, y sus valores de ITE y GRA. En la tabla se señala el valor de chi cuadrada de cada par y su correspondiente coeficiente phi, obtenidos según las ecuaciones (5) y (6). Un valor alto de phi significa un alto grado de asociación entre reactivos: una alta proporción de alumnos que contestaron ambos reactivos correctamente o ambos incorrectamente.

Se puede apreciar que la mayoría de los coeficientes son significativos, lo que prueba que los reactivos que tienen un mismo arreglo producen ejecuciones similares.

Los pocos casos en que phi no es significativo no deben interpretarse necesariamente como una baja asociación entre reactivos, sino como un efecto de distorisión en el cálculo

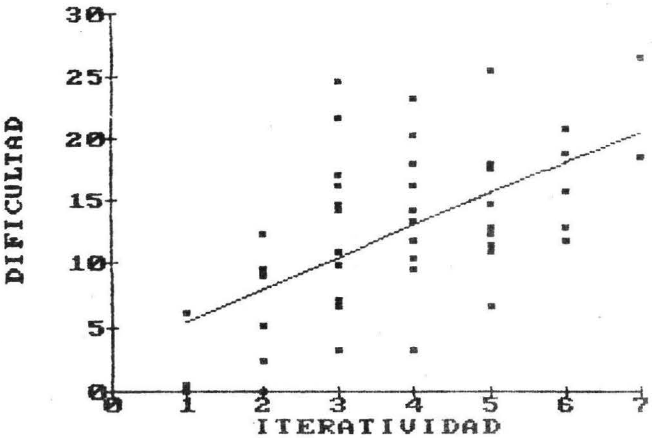


FIGURA 8. RELACIÓN ENTRE LA ITERATIVIDAD Y LA DIFICULTAD DE LOS REACTIVOS.

		Reactivo j	
		0	1
Reactivo i	0	A	B
	1	C	D

FIGURA 9. TABLA DE DOBLE ENTRADA ELABORADA PARA CADA PAR DE REACTIVOS FORMALMENTE EQUIVALENTES.

Tabla 9

VALORES DE ASOCIACION OBTENIDOS
EN PARES DE REACTIVOS FORMALMENTE EQUIVALENTES

ITE	GRA MUL	Reactivos i	Reactivos j	Dificultad DIF(i)	Dificultad DIF(j)	Chi cuadrada	phi
1	1	11	20	0.0	0.0	.042	.01457 +
2	1	23	27	9.0	9.95	20.934	.31498***
3	1	13	5	3.31	3.78	.285	.03677 +
2	2	1	29	2.36	6.16	1.696	.08967 +
3	2	65	22	14.69	10.42	9.203	.20885**
4	2	8	17	20.37	29.38	21.521	.31937***
5	2	50	43	6.63	16.58	.253	.03469 +
6	2	2	36	16.58	12.32	13.356	.25160***
3	3	26	48	16.11	19.90	3.939	.13663*
4	3	46	64	9.47	7.10	1.734	.09066 +
5	3	41	45	12.79	17.53	14.314	.26047***
6	3	39	52	11.84	28.90	13.876	.25644***
7	3	69	67	18.48	17.53	27.106	.35836***
4	4	40	47	20.37	14.69	10.406	.22208***
5	4	58	63	18.00	20.85	12.683	.24517***
6	4	61	3	18.95	15.63	22.199	.32436***

+ Valores no significativos

* $p < .05$

** $p < .025$

*** $p < .005$

de chi cuadrada: como puede comprobarse en la ecuación (6), el cálculo de chi cuadrada emplea en el numerador el producto de A por D. Si ninguno o pocos alumnos resuelven incorrectamente dos reactivos, el valor resultante de A es cero o cercano a cero, lo que disminuye el valor del numerador y, consecuentemente, de chi cuadrada. En la Tabla 9 puede constatarse que los pares de reactivos con un valor de phi no significativo son aquéllos que produjeron un valor muy bajo de DIF. Por ejemplo, los reactivos 11 y 20 produjeron un valor DIF de 0.0 (ningún alumno lo resolvió mal).

Por consiguiente, se acepta la hipótesis de que los reactivos con un mismo arreglo son resueltos ambos correctamente o ambos incorrectamente por una proporción significativa de alumnos.

DISCUSION

La aceptación parcial de la hipótesis 1 conduce a concluir que la predicción de la dificultad de un reactivo es más simple de lo que inicialmente se creyó: la dificultad de un reactivo es una función lineal directa de su iteratividad; el grado múltiple no ejerce influencia significativa. Con fines prácticos, esto puede resumirse en el siguiente principio: ***la iteratividad de un reactivo predice en buena medida su dificultad.***

Tal principio puede tener una enorme utilidad práctica, pues abre la posibilidad de construir exámenes con cierta objetividad, aun cuando no se cuente con información sobre la dificultad de los reactivos obtenida en aplicaciones previas.

Conviene reflexionar por qué, siendo ambas, variables relacionadas con la complejidad estructural de un problema, sólo la iteratividad, y no también el grado múltiple, influye en la dificultad de los reactivos. Una posible razón es que el valor máximo posible del grado múltiple es finito, mientras que el valor máximo de la iteratividad puede ser infinito. Un procedimiento cerrado como el que se empleó en esta investigación puede resolver, en teoría, problemas de ITE = 10000, 100000, o cualquier magnitud imaginable: tal posibilidad depende del código de representación, no de las relaciones entre los elementos del procedimiento. Sin importar cuán grande sea su iteratividad, un problema puede ser resuelto empleando las mismas variables lógicas y las mismas trayectorias que especifica el procedimiento. La iteratividad puede aumentar indefinidamente; el grado múltiple está limitado a un máximo valor posible.

Puede ser que la influencia del grado múltiple sobre la dificultad no resulte estadísticamente significativa por ser una variable que no tenga una suficiente cantidad de valores que "compita" con la cantidad de valores de iteratividad.

Para poner a prueba esta idea sería necesario realizar una investigación similar a ésta, en la que el procedimiento empleado tuviera una cantidad suficientemente grande de trayectorias posibles, de modo que pudiera haber valores altos de grado múltiple. No es difícil enconntrar un procedimiento con tales características. De hecho, en un trabajo anterior (Solano-Flores, en prensa) se presenta un algoritmo para la denominación de compuestos orgánicos de cadena abierta, usual en los cursos de química de nivel medio superior, que podría servir a tales fines.

Por otra parte, puede aceptarse la hipótesis 2, que establece que los reactivos con un mismo arreglo son resueltos ambos correctamente o ambos incorrectamente por una proporción significativa de alumnos. El hallazgo puede enunciarse mediante el siguiente principio: ***los reactivos formalmente equivalentes son resueltòs de manera equivalente.***

En otras palabras, los reactivos con el mismo arreglo múltiple, con las mismas propiedades estructurales, producen ejecuciones similares. El principio es de gran utilidad para agrupar a los reactivos según las propiedades formales que compartan; justifica también que la información sobre el arreglo múltiple sea incluida, junto con la dificultad o el índice de discriminación, en los bancos de reactivos; y aporta criterios formales para la construcción de bancos de reactivos y para la elaboración de exámenes paralelos.

Al menos desde hace veinte años, algunos autores han expresado su descontento porque el trabajo de elaboración de exámenes se efectúe de manera intuitiva. Tal deficiencia ha sido atribuida a la escasa atención otorgada al contenido de los reactivos (Guttman, 1969) y a la carencia de una base científica para su redacción (Bormouth, 1970). Wesman (1971) ha comentado que el trabajo de elaboración de exámenes tiene más de artístico que de científico.

Aunque se han propuesto formas alternativas para elaborar reactivos (Hively, 1974; Roid y Halladyna, 1982), los esfuerzos no han sido suficientes. Bejar (1985) ha señalado que para que la elaboración de reactivos pueda realizarse de manera científica, es necesario contar con medios para identificar las propiedades formales de los reactivos. De acuerdo con Embretson (1985), el trabajo de elaboración de exámenes ha estado limitado por la tradición de los estudios correlacionales en que se miden las diferencias individuales. Sólo recientemente han comenzado a ser incorporados los métodos provenientes de la psicología experimental al trabajo de elaboración de exámenes. Es posible concebir a quien elabora exámenes como alguien que, además de buscar datos de naturaleza estadística, controla y manipula variables que definen a las propiedades formales de los reactivos. El concepto de arreglo múltiple constituye la clave para definir e identificar esas propiedades formales. Por su parte, la iteratividad puede ser una de las variables que pueden ser controladas y manipuladas al diseñar exámenes.

C O N C L U S I O N E S

Además de señalar la utilidad práctica del DLE para elaborar exámenes y predecir la dificultad de los reactivos y la ejecución de los alumnos, conviene conclur este trabajo examinando las líneas de desarrollo que pueden derivarse de la técnica, y las aportaciones que ésta puede hacer a la investigación en áreas como la solución de problemas y la enseñanza. También conviene señalar las limitaciones de esta investigación y las dificultades que el DLE puede llegar a enfrentar antes de ser aceptado como técnica en el trabajo cotidiano de los especialistas o las personas relacionadas con la evaluación del aprendizaje.

LINEAS DE DESARROLLO Y APORTACIONES

Elaboración de modelos de solución de *problemas*

Existen antecedentes de investigaciones en las que se ha empleado un algoritmo como criterio para evaluar el aprendizaje. Hoc (1977) por ejemplo, evaluó el aprendizaje de un lenguaje de programación de una manera muy original: planteó un problema a sus sujetos y les pidió elaborar un algoritmo para su solución. El aprendizaje de los sujetos fue evaluado por la semejanza que sus algoritmos tenían con un algoritmo que el propio Hoc había elaborado.

Aunque no toma en cuenta la imposibilidad teórica de tener la certeza de que un algoritmo es el mejor (Lewis y Papadimitrou, 1976), tal forma de evaluación constituye un esfuerzo por determinar la estrategia seguida por los sujetos al resolver un problema. Sin embargo, sólo se ocupa de determinar la ejecución con base en un modelo de solución correcta, y no se detiene a analizar la naturaleza de las soluciones incorrectas.

Una línea diferente de investigación es la que postula al algoritmo como modelo, ya no de la ejecución ideal o correcta, sino de la estrategia seguida por el individuo al resolver una clase de problemas.

Siegler (1986) ha discutido el hecho sorprendente de que al resolver problemas matemáticos, los niños no necesariamente emplean los algoritmos que les enseñan en la escuela. Investigando al respecto, Groen y Parkman (1972) lograron desarrollar un modelo de solución de problemas de suma capaz de predecir la ejecución de niños de escuela primaria. Posteriormente, Ashcraft (1982) propuso un modelo alternativo. El modelo de Groen y Parkman predice mejor la ejecución de los niños antes del tercer año de primaria, mientras que el modelo de Ashcraft se ajusta mejor a la ejecución de los niños de cuarto año de primaria en adelante. Adicionalmente, ambos modelos predicen igualmente bien la ejecución de niños de tercer año. El hecho de que cada modelo se ajuste más a la ejecución de niños de ciertas edades, y de que ambos coincidan en una edad intermedia es un indicio de que, durante su desarrollo, los niños desarrollan diferentes estrategias para resolver problemas.

Para los propósitos de esta discusión, lo interesante de todo esto es que la precisión del modelo puede determinarse según su capacidad predictiva. Utilizando la misma lógica, varios algoritmos para un mismo tipo de problema pueden ser puestos a prueba: el que mejor prediga la ejecución de los sujetos podrá ser considerado como el mejor modelo.

El hecho de que en esta investigación se encontrara una fuerte asociación entre reactivos formalmente equivalentes, aun cuando éstos tuvieran un aspecto muy diferente (por ejemplo: **CCCLII** y **XXXVIII**), hace suponer que los sujetos tratan de resolver los problemas que se les presentan basándose principalmente en las relaciones entre los elementos del procedimiento, más que en otros factores como el código de representación. Esto parece cierto no sólamente cuando los sujetos resuelven correctamente los problemas, sino también cuando se equivocan. Ciertamente, se presentaron errores como el resolver **LXII** de la siguiente manera: **501011**, que implica basarse exclusivamente en el código de representación sin integrar a los elementos del problema de acuerdo con el procedimiento. Pero también fueron muy comunes los errores consitentes en que los elementos están correctamente relacionados, de acuerdo con el procedimiento, pero el código de representación es mal empleado. Por ejemplo, muchos de los alumnos que resolvieron mal el problema **LXII** (cuya solución correcta es **62**), lo resolvieron así: **602** (es decir, como si el problema fuera **DCII**). Debido a que **LXII** y **DCII** son formalmente equivalentes, puede decirse que una solución incorrecta como **602** es formalmente equivalente a la solución correcta: sorprendentemente, es una solución estructuralmente correcta.

Aunque hechas de manera muy incidental cuando se calificaron los exámenes de los sujetos, las observaciones anteriormente mencionadas dejan ver la posibilidad de concebir a los algoritmos no sólo como modelos de ejecución ideal, sino también como modelos de solución incorrecta de problemas. Esto puede ser útil para la evaluación diagnóstica de habilidades académicas (véase, por ejemplo, Enright, 1983). Si se elabora sistemáticamente un conjunto de algoritmos para un mismo tipo de problema (por ejemplo, la multiplicación de dos cantidades), cada uno de los cuales conduzca a soluciones incorrectas, se puede llegar a elaborar una clasificación de los errores posibles. De esta manera, la solución incorrecta ante un tipo de problema puede ser descrita por el algoritmo que conduzca a esa solución incorrecta.

Elaboración de reactivos de opción múltiple

Una ventaja de contar con tales algoritmos del error es la posibilidad de perfeccionar la elaboración de reactivos de opción múltiple. Aunque en este trabajo sólo fueron descritos los principios del DLE aplicables a la determinación del contenido de los reactivos, debe mencionarse que existe una técnica complementaria mediante la cual es posible elaborar sistemáticamente los distractores de reactivos de opción múltiple que evalúan el aprendizaje de procedimientos (Solano-Flores, en prensa).

Mediante tal tècnica, la opción correcta de un problema determinado se obtiene siguiendo la trayectoria correcta de acciones prescritas por el procedimiento; cada distractor se obtiene produciendo, en alguna iteración, cierta alteración en la trayectoria correcta de acciones, y resolviendo el problema de acuerdo con el algoritmo alterado. Las alteraciones, producidas de manera deliberado, no son sin ***manipulaciones de error***, que pueden ser de tres tipos: ***omisión de operación*** ("saltarse" o dejar de hacer lo que prescribe una operación), ***inversión de operación*** (hacer lo contrario de lo que especifica una operación), e ***inversión de decisión*** (invertir el sentido de la pregunta: contestar **SI** en vez de **NO**, o **NO** en vez de **SI**). De esta manera, los distractores son producidos sistemàticamente, cumpliendo con los requisitos de plausibilidad y verosimilitud señalados por autores como Thorndike y Hagen (1977) y Conoley y O'Neil (1979).

Secuenciación de la enseñanza de procedimientos

Una posible aportación del DLE consiste en la aplicación de sus principios analìticos en la determinación de secuencias de enseñanza. Varios autores han descrito tècnicas para determinar una secuencia óptima de enseñanza. Algunos se basan esencialmente en recursos gráficos (Morganov, 1966; D'hainaut, 1971; Castañeda, 1974; Le Xuan y Chassain, 1975; Huerta, 1978), mientras que otros emplean operaciones lógicas y matemáticas (Warfield, 1973; Solano-Flores, 1981, 1983). Pero todos se ocupan del problema de la secuenciación: las relaciones entre los elementos componentes de cualquier conocimiento son complejas (hay ramificaciones, circuitos y ciclos); en tanto, la enseñanza del conocimiento debe tener una estructura lineal en el tiempo. Con menor o mayor efectividad, dichas técnicas se basan en describir y analizar a la estructura de conocimiento y deducir, mediante ciertos principios lógicos, una estructura lineal óptima: una secuencia de enseñanza que favorezca al aprendizaje.

Por su nivel analítico, las técnicas mencionadas son más útiles para la enseñanza de conocimiento declarativo. Los ciclos, que representan un problema para la secuenciación de la enseñanza, resultan indispensables en cualquier procedimiento cerrado. Con propósitos de instrucción, los arreglos obtenidos como resultado de la segunda fase del DLE, análisis de variables, pueden ser considerados como una descripción sintética de todas las subclases de problema que deben ser considerados al enseñar un procedimiento. Sobre todo, pueden aportar información que contribuya a decidir el orden en que ha de secuenciarse la enseñanza de distintas subclases de problema. Por ejemplo, la enseñanza de un procedimiento puede iniciarse empleando como ejemplos o casos a los problemas que impliquen pocas decisiones què tomar (aquéllos cuyos arreglos tienen muchos *ceros* y pocos *unos* positivos o negativos) y culminar con problemas cuya solución requiera tomar muchas decisiones (aquéllos cuyos arreglos tienen pocos *ceros* y muchos *unos* positivos o negativos).

La tercera fase de la técnica tiene también algo qué ofrecer si se le emplea con fines de enseñanza y no de evaluación. El análisis lógico de reactivos (de problemas, en este caso) puede contribuir a que en un contenido dado, los problemas sean incluidos oportunamente de acuerdo con su iteratividad. Según se ha visto en este trabajo, tal variable es determinante de la dificultad de los problemas.

Estrategias de enseñanza de procedimientos

Aplicar los principios analíticos del DLE con propósitos de enseñanza ofrece una posibilidad adicional: la de transformar la estrategia de enseñanza de un procedimiento. Cuando un procedimiento es cerrado, puede suceder que la mayoría de los problemas que resuelve sean de iteratividad mayor que 1, y que pocos problemas tengan tal iteratividad. Según su arreglo, los problemas de ITE = 1, del procedimiento empleado en esta investigación son los siguientes:

a = (1 -1 -1): **I, X, C, M;**

b = (1 1 0): **II, III, XX, XXX, CC, CCC, MM, MMM;**

e = (1 -1 1): **IV, IX, XL, XC, CD, CM;**

f = (-1 0 0): V, L, D.

Como se puede ver, se trata de sólo 21 problemas con ITE = 1: cantidad minúscula, si se le compara con el total de 3,999 problemas que el procedimiento, limitado hasta el símbolo **M**, puede resolver. Los 3,978 con ITE mayor que 1, no son sino combinaciones de los 21 problemas enlistados arriba. De acuerdo con ello, un problema podría resolverse de dos maneras: mediante un algoritmo como el de la Figura 2, o segmentándolo en bloques de ITE = 1. Mediante la segunda estrategia, el problema **MCMLXXXIX** se resolvería identificando los bloques: **M**, **CM**, **L**, **XXX**, y **IX**.

Es posible que, si se le enseña a memorizar como bloques indivisibles a todos los problemas con ITE = 1, una persona pueda resolver problemas tan complejos como **MCMLXXXIX**, aun sin conocer el algoritmo. La enseñanza de tal estrategia implica, por ejemplo, que **IX** se enseñe como **9** y no como la sustracción **10 - 1 = 9**.

Así, parece que la enseñanza "mediante bloques" de un procedimiento, puede ser más económica. En el caso de la lectura de números romanos parece ser aun más económica, si se considera que, de los 21 bloques arriba enlistados, siete tienen que ser aprendidos como código básico para poder operar un algoritmo como el de la Figura 2.

Por supuesto, será necesaria una investigación sistemática para determinar si realmente es mejor o más económico enseñar un procedimiento por bloques, o en qué medida lo que aquí se ha dicho es generalizable a otros procedimientos cerrados. Mientras tanto, es preciso señalar tres hechos importantes. El primero es que la iteratividad de un problema es igual a la cantidad de bloques que lo componen, de modo que lo dicho en esta sección es compatible con los planteamientos que dieron origen a este trabajo.

El segundo hecho es que los bloques de un procedimiento no se pueden determinar "a primera vista": se requiere para ello efectuar el análisis de variables. Dicho de otra manera: la enseñanza de un procedimiento por bloques no podría efectuarse basándose sólamente en el sentido común.

El tercer hecho es muy interesante: al obtener una lista de bloques, el conocimiento tiende a dejar de ser conocimiento procedural para transformarse en conocimiento de tipo declarativo. Para que esto quede claro, analícese la Figura 10, que presenta el diagrama del procedimiento de lectura de números romanos, cuando los problemas son resueltos por bloques. Mientras que el algoritmo de la Figura 2 tiene tres decisiones de reconocimiento, el algoritmo de la Figura 10 no tiene ninguna: no hay variables lógicas qué identificar. A cambio, su código es más amplio: contiene 21 bloques que hay que memorizar, y no siete

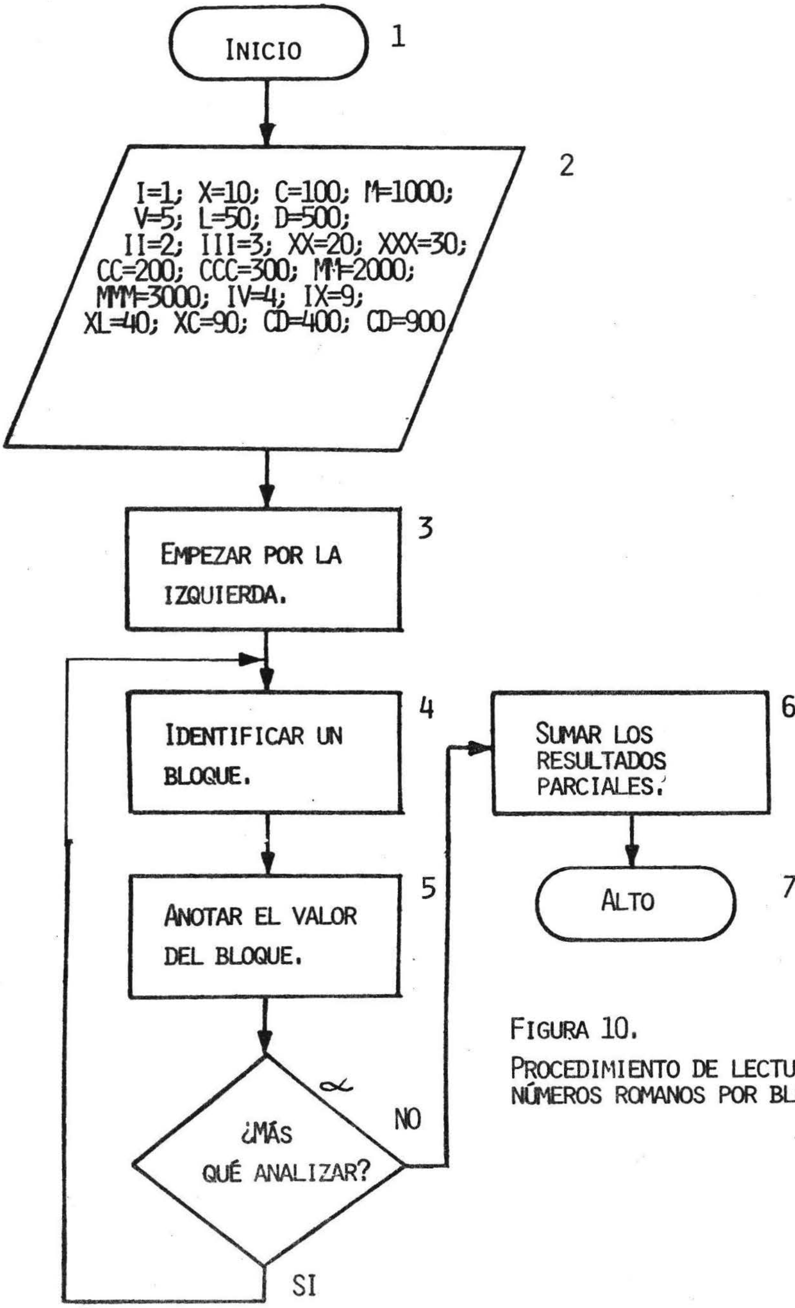


FIGURA 10.

PROCEDIMIENTO DE LECTURA DE NÚMEROS ROMANOS POR BLOQUES.

símbolos, como en el algoritmo de la Figura 2. Qué tan eficiente es la enseñanza de un procedimiento mediante bloques, depende posiblemente de una relación costo-beneficio: menos decisiones de reconocimiento que tomar significan más bloques que memorizar; menos bloques que memorizar significan más decisiones de reconocimiento.

Al principio de este trabajo se mencionó que el experto sigue una estrategia prodroma: identifica y resuelve con rapidez el problema que tiene enfrente, y no considera secuencias alternativas de acción. Parece razonable pensar que la estrategia seguida por un experto al resolver un problema se asemeja más al tipo de algoritmo de la Figura 10 que al de la Figura 2. Esto tiene el apoyo de una conclusión a la que llega Anderson (1985), en relación con los expertos: el experto adquiere su pericia por la multitud de problemas a las que ha sido expuesto anteriormente, gracias a la información que acumula. De acuerdo con esto, la pericia podría estar relacionada con los bloques que una persona es capaz de manejar.

LIMITACIONES

Conocimiento procedural

En cuanto a las limitaciones de esta investigación, debe mencionarse inicialmente el hecho de que el DLE esté restringido a la evaluación del aprendizaje de conocimiento procedural. La extensión del DLE a la evaluación del aprendizaje de conocimiento declarativo dependerá de la medida en que sus propiedades descriptivas y analíticas pueden ser aprovechadas para representar las relaciones entre los elementos constituyentes de conceptos. Estructuralmente, un concepto puede ser más complejo que un procedimiento entero: cualquiera que éste último sea, las relaciones entre sus elementos son exclusivamente de secuencia. En cambio, en un concepto pueden participar relaciones de muy diversa índole: secuencia, jerarquía, inclusividad, disyunción, conjunción, etcétera. No es difícil imaginar que para extender las aplicaciones del DLE a la evaluación del aprendizaje de conocimiento declarativo, lleguen a ser empleadas las variables lógicas para describir la presencia o ausencia de atributos y la noción de arreglo múltiple para describir los distintos tipos de relación entre elementos. Pero todo ello requerirá un intenso trabajo de desarrollo teórico cuya naturaleza no podría ser anticipada en este espacio.

Generalización de los resultados

Otra limitación de este trabajo se deriva del empleo de un procedimiento específico. Este es un problema de validez externa inherente a cualquier investigación como la presente. Evidentemente, se requiere determinar en qué medida los resultados encontrados son generalizables a otros procedimientos. Sería interesante explorar, por ejemplo, la capacidad de la iteratividad para predecir la dificultad de los reactivos de otros procedimientos matemáticos y, más aún, de los reactivos de procedimientos de naturaleza no matemática.

También será interesante determinar cuál es el mejor predictor de la dificultad de los reactivos, o cómo se puede concebir a los reactivos formalmente equivalentes, cuando un procedimiento tiene propiedades estructurales diferentes a las del procedimiento empleado en esta investigación. La solución de problemas por aproximaciones y, por ende, la necesidad de que la solución de un problema ocupe más de una iteración, es característica de los procedimientos cerrados. Habiéndose encontrado en este trabajo que un buen predictor de la dificultad de los reactivos es la iteratividad, conviene preguntarse qué variable de orden estructural puede ser responsable de la dificultad de los reactivos cuando el procedimiento no es cerrado. Es el caso de procedimientos que no contienen una decisión candado (Solano-Flores, en prensa) como la decisión α en el procedimiento de lectura de números romanos, que remita a partes del procedimiento anteriormente recorridas.

Dificultades para la adopción de la técnica

Otras limitaciones de esta investigación y del DLE son de orden práctico: la primera es que la técnica puede ser dificil de entender para quienes no tienen suficiente familiaridad con ciertas nociones de lógica o de matemáticas; la segunda, que la técnica puede ser demasiado laboriosa.

En el fondo, se trata de limitaciones derivadas de la resistencia al cambio. Horabin y Lewis (1978) señalan ventajas incontrovertibles del empleo de algoritmos como medio instruccional (por ejemplo: que éstos presentan instrucciones de manera clara y breve, que ocupan poco espacio de impresión y que exigen que el usuario sólamente emplee o lea de ellos lo que requiere para su problema especifico). Los mismos autores reconocen, sin embargo, que existe una seria dificultad en el ámbito laboral y, en general, entre los posibles usuarios, para generar y emplear algoritmos como medio usual de consulta. Si, a pesar de sus ventajas como medio instruccional, los algoritmos son aceptados con dificultad entre el común de la gente, cabe suponer que aún más dificil será incorporar a la práctica cotidiana de las instituciones educativas o relacionadas con la evaluación del aprendizaje, una técnica como el DLE que, además, emplea abstracciones como las de variable lógica, arreglo y arreglo múltiple. Aprender la técnica puede resultar dificil para un profesor común que, a diferencia de un ingeniero, un analista de sistemas o un contador, está poco familiarizado con grafos, diagramas de flujo y variables lógicas.

En realidad, no hay razón para pensar que el DLE deba ser necesariamente accesible a todo mundo para que se justifique su adopción. Tal vez ni siquiera deba pensarse en que la técnica sea aplicada por el profesor común que desea elaborar sus propios exámenes: a éste le resultaría poco costeable invertir esfuerzo y tiempo en elaborar un examen de acuerdo con los principios del DLE, si puede elaborar un examen de manera intuitiva. Quizá valga más la pena pensar en que un grupo de especialistas (o de personas que se tomen la molestia de aprender la técnica y aplicarla sistemáticamente) identifiquen todos los procedimientos que forman parte de los contenidos de un cierto nivel educativo (por ejemplo: suma, resta, multiplicación, división, raíz cuadrada, fracciones, etcétera, para la enseñanza primaria) y elaboren catálogos de algoritmos y bancos de reactivos elaborados según la técnica descrita en este trabajo, que puedan ser empleados aun por los no especialistas.

Debe considerarse el costo que representa aprender una técnica como el DLE y el beneficio que puede significar su uso regular. Para un procedimiento determinado, las dos primeras fases de la técnica (formalización y análisis de variables), sólo deben llevarse a cabo una vez. El árbol expansivo y el listado de arreglos obtenidos en estas fases sirven para la creación y el análisis lógico de todos los reactivos que posteriormente se requieran. Sólo la tercera fase de la técnica (creación y análisis lógico de reactivos) debe ser aplicada cuando se requiera crear reactivos o determinar mediante criterios formales su dificultad o su equivalencia. Aun esta etapa puede no ser requerida si se cuenta con un adecuado banco de reactivos.

Según esta idea, un profesor, un equipo de profesores, el departamento de didáctica de una institución, o cualquier institución que por una u otra razón requiera evaluar periódicamente, póngase por caso, el aprendizaje del procedimiento para balancear ecuaciones químicas, tendrá necesidad de aplicar una sola vez las dos primeras fases de la técnica, para elaborar el algoritmo y efectuar el análisis de variables. Después de eso, cualquier examen que se elabore sólo requerirá crear y analizar reactivos. Y si ya existe un banco de reactivos, clasificados según sus propiedades formales, sólo se requerirá tomar los reactivos que satisfagan las necesidades de evaluación.

El hecho de que el DLE facilite y haga más objetivo el trabajo de elaboración de reactivos, es suficiente justificación para aprender la técnica, sobre todo si se considera que un curso de capacitación para su menejo, dirigido a pedagogos o psicólogos educativos, no debe consumir más de 20 horas de entrenamiento.

De cualquier manera, con la difusión de las computadoras y de ciertos conceptos provenientes de la informática, es posible que en un futuro breve, nociones como las de variable lógica o diagrama de flujo se asimilen con más facilidad en la cultura general de las personas. La investigación aplicada parece indicar que existe interés por hacer que tales elementos sean accesibles. Krohn (1983), por ejemplo, describe un trabajo en el que se determinó la mejor manera de ubicar en una página los componentes de un diagrama de flujo para facilitar su interpretación.

Pero más que cualquier argumento, lo que más puede facilitar la difusión del DLE entre especialistas en elaborar exámenes, será el desarrollo de paquetes de programas de cómputo que empleen los principios de la técnica. En virtud de las propiedades lógicas de la técnica, es posible imaginar las funciones de dichos programas: elaboración de algoritmos, identificación de errores lógicos en la elaboración de algoritmos, determinación de las variables lógicas implicadas en un procedimiento, elaboración de árboles expansivos, determinación de la iteratividad de los reactivos, identificación de los arreglos múltiples de los reactivos, construcción de bancos de reactivos según criterios formales, etcétera.

En tanto se emprenden las investigaciones necesarias para esclarecer las dudas planteadas y para difundir el empleo de la técnica, puede concluirse que su corrección formal y las ventajas prácticas que ofrece, son buenas razones para que se considere la incorporación del DLE a la tecnología moderna para la elaboración de exámenes.

R E F E R E N C I A S

Acosta, M. A. y Stockton, F. (1979). Validez y confiabilidad en las pruebas objetivas de rendimiento escolar. En Fernando García (Ed.), *Paquete de autoenseñanza de evaluación del rendimiento escolar*. México: UNAM, CISE.

Anderson, J. R.(1976). *Language, memory, and thought.* Hillsdale, Nueva Jersey: Lawrence Erlbaum Associates, Publishers.

Anderson, J. R. (1983). *The architecture of cognition.* Cambridge, Massachusetts: Harvard University Press.

Anderson, J. R. (1984). Cognitive Psychology. *Human Factors*, 1984, **23**, 1-11.

Anderson, J. R. (1985). *Cognitive psychology and its implications*. Nueva York: W. H. Freeman and Company.

Ashcraft, M. H. (1982). The development of mental arithmetic: A chronometric approach. *Development Review*, 2, 213-236.

Bejar, I. I. (1985). Speculations on the future of test design. En S. E. Embretson (Ed.) *Test design: developments in psychology and psychometrics.* Nueva York: Academic Press.

Bormouth, J. R. (1970). *On the theory of achievement test items*. Chicago, Illinois: University of Chicago Press.

Castañeda, M. (1971). *Métodos de análisis para la enseñanza de un contenido.* Tesis profesional. México: U.N.A.M., Facultad de Psicología.

Cedeño, M. L. y Ruiz-Primo, M. A. (1982). *Una estrategia para evaluar habilidades metodológico-conceptuales.* Tesis profesional. México, UNAM, Facultad de Psicología.

Coffman, W. E. (1971). Essay examinations. En R. L. Thorndike (Ed.), *Educational measurement*. Washington, D. C.: American Council of Education.

Cohen, J. y Cohen, P. (1975). *Apllied multiple regression/correlational analysis for the behavioral sciences*. Hillsdale, Nueva Jersey: Lawrence Erlbaum Associates.

Conoley, J. C. y O'Neil, H. P. Jr. (1979). A primer for developing test items. En H. F. O'Neil, Jr. (Ed.), *Procedures for instructional systems development*. Nueva York: Academic Press.

D'hainaut, L. (1971). *L'enseignement des concepts scientifiques a l'aide de cours programmes*. Bélgica: Universidad Libre de Bruselas.

Ebel, R. (1972). *Essentials of educational measurement*. Englewood Cliffs, Nueva Jersey: Prentice-Hall.

Embretson, S.E. (1985). Introduction to the problem of test design. En S. E. Embretson (Ed.) *Test design: developments in psychology and psychometrics*. Nueva York: Academic Press.

Enright, B. E. (1983). *ENRIGHT diagnostic inventory of basic arithmetic skills*. North Billerica, Massachusetts: Curriculum Associates, Inc.

Farina, M. V. (1970). *Flowcharting*. Englewood Cliffs, Nueva Jersey: Prentice-Hall.

Groen, G. J. y Parkman, J. M. (1982). A chronometric analysis of simple addition. *Psychological Review, 79*, 329-343.

Gronlund, N. E. (1971). *Measurement and evaluation in teaching*. Nueva York: Macmillan.

Guttman, L. (1969). Integration of test design and analysis. En *Proceedings of the 1969 invitational conference on testing problems*. Princeton, Nueva Jersey: Eductional Testing Service.

Harary, F. (1969). *Graph theory*. Reading, Massachusetts: Addison-Wesley.

Hively, W. (1974). Introduction to domain referenced testing. *Educational Technology, 14*, 5-9.

Hoc, J. M. (1977). Role of mental representation in learning a programming language. *International Journal of Man-Machine Studies*, **9**, 87-105.

Horabin, I. y Lewis, B. (1978). *Algorithms*. Englewood Cliffs, Nueva Jersey: Educational Technology Publications.

Huerta, J. (1978). *Organización lògica de las experiencias de aprendizaje*. Mèxico: Trillas.

Jackson, D. y Gonzàlez, M. D. (1979). *Introducción a la teoría de gráficas en el campo de la educación*. Mèxico: Asociación Nacional de Universidades e Institutos de Enseñanza Superior.

Krohn, G. S, (1983). Flowcharts used for procedural instructions. *Human Factors*, **25**(5), 573-561.

Landa, L. N. (1972). *Cibernètica y pedagogía*. Barcelona: Labor.

Landa, L. N. (1974). *Algorithmization in learning and instruction*. Englewood Cliffs, Nueva Jersey.

Le Xuan, M. y Chassain, J. C. (1975). *Analyce de contenu-analyce comportamentale*. Laos: Vientiane.

Levin, J. (1979). *Fundamentos de estadística en la investigación social*. Mèxico: Harla.

Lewicki, P., Hill, T. y Bizot, E. (1988). Acquisition of procedural knowledge about a pattern of stimuli that cannot be articulated. *Cognitive Psychology*, **20**, 24-37.

Lewis, H. R. y Papadimitrou, C. H. (1978). La eficiencia de los algoritmos. *Investigación y Ciencia: edición en español de Scientific American*, **18**. 78-91.

Magnusson, D. (1976). *Teoría de los tests*. Mèxico: Trillas.

Martin, J. y McClure, C. (1985). *Diagramming techniques for analysts and programmers*. Englewood Cliffs, Nueva Jersey.

Millman, J. (1980). Computer-based item generation. En R. A. Berk (Ed.), *Criterion-referenced measurement: the state of the art*. Baltimore: The Johns Hopkins University Press.

Morganov, I. B. (1966). L'utilisation des graphes dans l'elaboration de programmes. *Enseignement programeè*, núm. 1.

Nie, N. H., Hull, C. H., Jenkins, J. G., Steinbrenner, K. y Bent, D. H. (1975). *Statistical Package for the social sciences*. (Segunda edición). Nueva York: McGraw-HIll.

Nunnally, J. C. (1964). *Educational measurement and evaluation*. Nueva York: McGraw-Hill, 1964.

Nunnally, J. C. (1967). *Psychometric theory*. Nueva York: McGraw-Hill.

Payne, D. A. (1968). *The specification and measurement of learning outcomes*. Lexington, Massachusetts: Xerox.

Priestley, M. (1982). *Performance assessement in education and training: alternative techniques*. Englewood Cliffs, Nueva Jersey: Educational Technology Publications.

Roid, G. H. (1979). The technology of test-item writing. En H. F. O'Neil, Jr. (Ed.), *Procedures for instructional systems development*. Nueva York: Academic Press.

Roid, G. H. y Haladyna, T. M. (1978). A comparison of objective-based and modified-Bormouth item writing techniques. *Educational and Psychologycal Measurement*, **35**, 19-28.

Roid, G. H. y Haladyna, T. M. (1982). *A technology for test-item writing*. Nueva York: Academic Press.

Salazar-Resines, J. (1979). *Enfoque de sistemas en la educaciòn: teorìa de gràficas*. Mèxico: Limusa.

Schriber, T. J. (1969). *Fundamentals of flowcharting*. Nueva York: Wiley.

Schyfter, G., Stockton, F., Trejo, M. A. y Garcìa, F. (1979) Elaboraciòn de pruebas objetivas: redaciòn de reactivos para evaluar el logro de objetivos de aprendizaje. en F. Garcìa (Ed.), *Paquete de autoenseñanza de evaluaciòn del aprovechamiento escolar*. Mèxico: UNAM, CISE.

Siegler, R. S. (1986). *Children's thinking*. Englewwod Cliffs, Nueva Jersey: Prentice-Hall.

Solano-Flores, G. (1981). Tècnica algebraica para la realizaciòn de anàlisis de contenido. *Enseñanza e Investigaciòn en Psicologìa*, 2 (14), 307-324.

Solano-Flores, G. (1982). El anàlisis estructural educativo: ventajas y aplicaciones. *Acta Psicològica Mexicana*, **1** (4), 9-21.

Solano-Flores, G. (1983). *Principios de anàlisis estructural educativo: metodologìa y tècnicas para la educaciòn*. Mèxico: Trillas.

Solano-Flores, G. (1985). *La formalización de algoritmos como base para la evaluación del dominio de procedimientos.* Investigación inédita. México: Facultad de Psicología, UNAM.

Solano-Flores, G. (En prensa). *Diseño lógico de exámenes.* México: Trillas.

Stern, N. B. (1975). *Flowcharting: a tool for understanding computer logic.* Nueva York: Wiley.

Thorndike, R. L. y Hagen, E. (1977). *Tests y técnicas de medición en psicología y educación.* México: Trillas.

Trakhtenbrot, B. A. (1963). *Algorithms and automatic computing machines.* Chicago: University of Chicago Press.

Warfield, J. N. (1973). On arranging elements of a hierarchy in graphic form. *IEEE Transactions on Systems, Man and Cybernetics, 3* (2), 121-132.

Wesman, A. G. (1971). Writing the test item. En R. L. Thorndike (Ed.) *Educational measurement.* Washington, D.C.: American Council of Education.

Wheatley, D. M. y Unwin, A. W. (1972). *The algorithm writer's guide.* Londres: Longman.

Wirth, N. (1976). *Algorithms + data structures = programs.* Englewood Cliffs: Nueva Jersey.