

EL ANÁLISIS COGNITIVO DE TAREAS, BASE PARA EL DISEÑO DE INSTRUMENTOS DE EVALUACIÓN EN EL APRENDIZAJE EN LÍNEA

Eduardo Peñalosa y Sandra Castañeda

Revista Iberoamericana de Evaluación Educativa 2009 - Volumen 2, Número 1

<http://www.rinace.net/riee/numeros/vol2-num1/art9.pdf>

Fecha de recepción: 30 de marzo de 2009

Fecha de comunicación de dictamen: 29 de abril de 2009

Fecha de aceptación: 29 de abril de 2009

Se estima que el aprendizaje en línea es un proceso activo y complejo, en el que el estudiante construye sus conocimientos con base en una estructura de significados previos, mediante la interacción con agentes humanos o materiales en entornos virtuales. Para el desempeño de estos procesos se ha sugerido que los ambientes de aprendizaje en línea estén centrados en: a) el estudiante como protagonista y agente; b) el conocimiento y la competencia; c) la evaluación, para ayudar a los estudiantes a tener una valoración de su propia comprensión, y d) la comunidad, que propician la colaboración con otros estudiantes y miembros de la misma (Bransford, Brown y Cockings, 2004; Anderson, 2004).

De entre las funciones enunciadas antes, en este trabajo interesa en especial la de evaluación del desempeño en contextos de aprendizaje en línea. Se describe la construcción de evaluaciones sensibles del aprendizaje, que permiten identificar la influencia de elementos del ambiente y del diseño instruccional. Para identificar el progreso de los procesos cognitivos e interactivos en entornos virtuales de aprendizaje es necesario contar con una estrategia sensible y válida de evaluación del desempeño, así como una serie de herramientas que permitan detectar cambios en la complejidad de las construcciones de conocimientos por parte de los estudiantes.

Existen tres etapas en la evaluación del aprendizaje presencial o en línea: 1) la que se realiza inicialmente de manera que tengamos una idea fiable del nivel de conocimientos y habilidades de los estudiantes (evaluación inicial); 2) la que se realiza durante la acción docente, a intervalos establecidos ya en la fase de diseño y planificación (evaluación formativa), y 3) la que se realiza al finalizar la acción docente (evaluación sumativa).

Dadas las características de la educación en línea, donde existen limitaciones en la comunicación impuestas por la naturaleza del medio, la evaluación continua del aprendizaje es esencial, por su función de retroalimentación al estudiante acerca de su desempeño. Quesada (2006) plantea que los atributos de una evaluación en línea deben ser: confiabilidad, validez, objetividad y autenticidad. La primera se refiere a la confianza que genera un instrumento para reflejar el nivel de logro del estudiante; la segunda se refiere a que el instrumento mida lo que realmente se pretende, y no otra cosa; la tercera se refiere a la neutralidad con la que se califica a los estudiantes, y la última se relaciona con los contenidos de la evaluación con la realidad práctica de los estudiantes. Quesada indica que si bien se ha escrito mucho sobre educación a distancia "en línea", pocas referencias se dedican al análisis de la evaluación en estos contextos.

En la investigación acerca del aprendizaje en línea, Peñalosa y Castañeda (2008a) citan alrededor de una decena de estudios en los que se analiza la eficacia de cursos en línea de Psicología, y las estrategias de evaluación son, fundamentalmente: 1) la realización de cuestionarios de conocimientos que se presentaban al finalizar los cursos, con un dudoso cumplimiento de criterios de confiabilidad, validez o autenticidad, aunque con cierta garantía de objetividad, inherente al medio en línea, y/o 2) la aplicación de entrevistas o inventarios que sondean la satisfacción del estudiante en el curso, lo cual implica considerar como iguales satisfacción y aprendizaje, dos aspectos no necesariamente equiparables, ya que la variable satisfacción hace mención más a efectos motivacionales que instruccionales.

La variedad de recursos de evaluación del aprendizaje en línea incluye pruebas objetivas, ensayos, proyectos, rúbricas, creación de mapas conceptuales o evaluación basada en portafolios de evidencias, por citar las más utilizadas. Una clasificación posible de los tipos de evaluación en la educación en línea incluye evaluaciones: a) automáticas, programadas para ser resueltas y calificadas de inmediato; b) elaborativas, en ellas el estudiante construye un producto como un ensayo, un mapa conceptual, un

proyecto, etc.; y c) colaborativas, en ellas los estudiantes son evaluados en situaciones de trabajo grupal (clasificación adaptada de Barberá, 2006).

Un paso crítico en el desarrollo de sistemas centrados en el estudiante consiste en tener información de calidad acerca de las competencias del estudiante en el dominio educativo. La tecnología incluye un conjunto de herramientas con potencialidad para optimizar la evaluación de los aprendizajes en entornos en línea. De hecho, en la modalidad educativa en línea, la evaluación debería ser una actividad continua, el eje vertebrador del aprendizaje (Barberá, 2006).

Sin embargo, existe coincidencia entre algunos autores acerca de que si bien en la puesta en marcha de proyectos de educación virtual se ha prestado atención a la pertinencia de los contenidos, al diseño instruccional, al uso óptimo de los recursos tecnológicos, por otra parte se ha dejado de lado la consideración minuciosa de la evaluación del aprendizaje que con ellos se logra (Quesada, 2006), y de hecho la evaluación se ha considerado un "punto débil" de la educación virtual (Barberá, 2006).

La infraestructura tecnológica podría permitir el registro flexible y exhaustivo de los datos de desempeño y el seguimiento de los progresos de los alumnos en su proceso de desarrollo cognitivo, de construcción de conocimientos y adquisición de habilidades. En estos contextos el sistema sigue la huella de lo que el estudiante hace, no omite partes de su desempeño y tampoco emite juicios parciales: tiene la programación de los componentes que se deben cumplir y registra los datos en consecuencia. Una evaluación tan minuciosa sería difícil de llevar a la práctica por parte de un profesor, y por su minuciosidad se podría prestar a omisiones. Para un curso determinado, se programa una vez y se utiliza para cientos de alumnos. Adicionalmente, se convierte en un laboratorio para identificar las mejores prácticas instruccionales y los efectos de los distintos tipos de esquemas interactivos.

A pesar de tal potencialidad, la evaluación del aprendizaje en línea es "algo pendiente de mejorar, algo sobre lo que hay que reflexionar con mayor profundidad, que hay que desarrollar de un modo más creativo y volver sobre ello invirtiendo más recursos pedagógicos y tecnológicos" (Barberá, 2006: 4).

De esta manera, tenemos que si bien la evaluación se considera parte fundamental del proceso de aprendizaje en línea, y aun cuando se reconoce un gran potencial de la tecnología para crear sistemas efectivos de evaluación, no existe un desarrollo relevante de aplicaciones tecnológicas para realizarla.

La propuesta de este trabajo se desarrolla en el dominio de un curso de psicología clínica teórica. Dada la necesidad de contar con este tipo de evaluaciones que brinden información acerca de los niveles de desempeño en los que impacta la intervención, en el presente estudio se reporta la metodología para construir y validar instrumentos de evaluación para la enseñanza en línea de contenidos de psicología clínica teórica, así como los resultados de dicha validación.

Dada la identificación de métodos instruccionales tradicionales, que conducen a un aprendizaje superficial o inerte, que permite a los estudiantes recordar o repetir lo aprendido, pero no utilizar el conocimiento nuevo como herramienta; esto es, no demostrar la capacidad de aplicación ni transferencia del mismo (Renkl, Mandl y Gruber, 1996; Schwartz, Brophy, Lin y Bransford, 1999), se hace necesario contar con herramientas diagnósticas que permitan considerar los diversos niveles de complejidad de los procesos subyacentes al desempeño, así como evaluar integralmente la progresión del estudiante en las diversas etapas y capas del aprendizaje. Esto es posible con ayuda de un análisis diferencial de operaciones cognitivas de complejidad creciente; esto es, con un *Análisis Cognitivo de Tareas (ACT)* (Castañeda, 1998a; 1998b; 2002; 2004a; 2004b; 2006; Chipman, Schragen y Shalin, 2000).

1. ANÁLISIS COGNITIVO DE TAREAS Y EVALUACIÓN DEL DESEMPEÑO

Para realizar un ACT, en primera instancia es necesario identificar las competencias que componen al dominio educativo. Podemos plantear que esta identificación de las competencias principales de un dominio constituye la macroestructura de lo que se desea evaluar (Castañeda, 2006). Cuando contamos con ella, se procede a realizar un ACT para identificar los microcomponentes que constituyen a cada competencia, e identificar gradientes de complejidad basados en las demandas de procesamiento generadas en la situación de evaluación. El procedimiento analiza la tarea mediante pasos en los que se identifican, en una secuencia progresiva de mayor detalle y precisión, los conocimientos, habilidades y las disposiciones asociados a cada paso. En este proceso es importante identificar la complejidad cognitiva (p. ej.: comprensión, aplicación, solución de problemas) y el tipo de modelo mental (p. ej.; conceptual, estructural, causal).

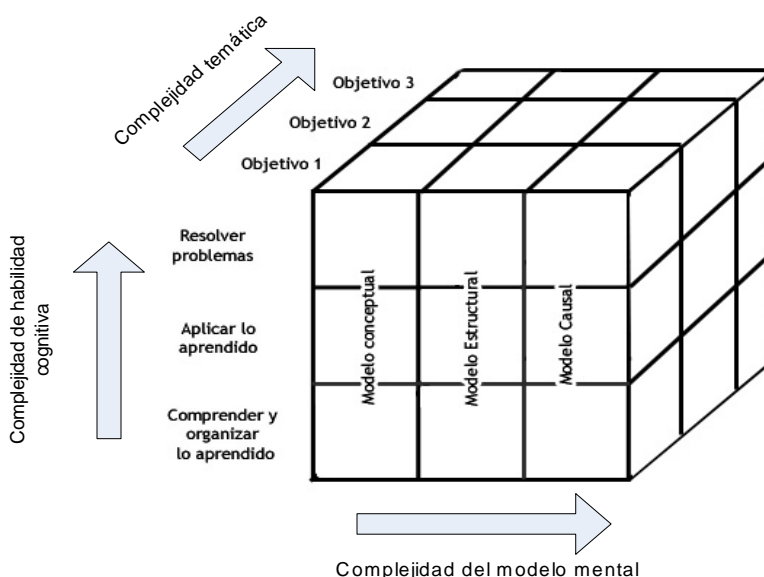
Con base en ambos análisis, se construye una estructura que identifica componentes de grano fino sin perder de vista las ligas de las competencias de las que forman parte. De esta forma se organiza un universo de medida tal que ha permitido diseñar situaciones de evaluación representativas de competencias críticas y sus componentes, en un esquema articulado y anidado de evaluación.

Como se expone arriba, en el proceso de análisis de componentes de las tareas es preciso clasificar los pasos del desarrollo de las competencias, y de esta forma se obtienen gradientes de complejidad de las competencias. Las dimensiones para realizar estas clasificaciones son:

1. Un continuo de complejidad en cuanto a las habilidades cognitivas requeridas en el dominio, que incluye tres categorías:
 - a) comprensión de los temas, que incluye el reconocimiento de información, la clasificación, el ordenamiento, el resumen;
 - b) aplicación del conocimiento y la habilidad, que incluyen operaciones como la traducción, el análisis, la extrapolación, la inferencia, la comparación o la aplicación de procedimientos, y,
 - c) solución de problemas, que incluye operaciones como corrección de errores, planificación de acciones, evaluación y toma de decisiones.
2. La dimensión de complejidad en cuanto a los modelos mentales construidos por los estudiantes. Las habilidades cognitivas representan categorías de operaciones que el estudiante puede aplicar en el dominio; los modelos mentales reflejan la integración de conocimiento que le permiten explicar la realidad de los fenómenos. La inclusión de los modelos mentales como un eje de diagnóstico responde a evidencias de que los estudiantes, en función de su nivel de pericia, podrían describir sólo conceptualmente el dominio, o podrían tener un conocimiento altamente estructurado acerca del mismo, por lo cual podría explicarlo. Los modelos mentales que se incluyen en el presente modelo son tres:
 - a) conceptuales, responden a la pregunta: "¿qué es esto?", describen el significado del fenómeno o tema, y la interrelación de los elementos que lo componen;
 - b) estructurales, que responden a la pregunta: "¿cómo está estructurado esto?", y describen cómo está organizado el campo conceptual en cuestión; y,

- c) causales, que responden a la pregunta: "¿cómo funciona esto?", y que describen cómo los principios se afectan entre sí y ayudan a interpretar procesos, dar explicaciones de eventos y realizar predicciones.
3. La dimensión de la complejidad de los temas de aprendizaje, determinada por las unidades de contenido como universo a evaluar aplicando diferentes modelos mentales con diferentes niveles de complejidad cognitiva. Esta dimensión se ordena de lo simple a lo complejo, en función de los objetivos de las unidades del programa de aprendizaje elegido.

FIGURA 1. MODELO DEL ANÁLISIS COGNITIVO DE TAREAS DE UN DOMINIO



La figura 1 muestra la estructura que resultaría de integrar las tres dimensiones relevantes del análisis. En ella cada espacio del cubo incluye una categoría de contenido que juntas conforman el universo del dominio educativo analizado. Esto significa que de cada casilla pueden generarse reactivos de evaluación, recursos y actividades de aprendizaje.

Por ejemplo, tomemos el dominio educativo de la evaluación en psicología clínica. La casilla inferior izquierda tendría contenidos del primer objetivo (digamos antecedentes teóricos de la evaluación psicológica), del nivel cognitivo de comprensión (implica identificar definiciones de evaluación, de evaluación como signo, como muestra), y del modelo mental conceptual (el estudiante debería encontrar instancias de tareas de evaluación psicológica entre varias disponibles).

En el mismo dominio, una casilla que implica mayor complejidad: imaginemos la superior derecha de la cara frontal del cubo. En ella seguimos en el primer objetivo (antecedentes teóricos de la evaluación psicológica), pero ahora el nivel cognitivo es el tercero: solución de problemas (por ejemplo, el estudiante identifica errores en casos resueltos), y el nivel de complejidad de su modelo mental debe ser causal (debe explicar la interrelación funcional entre los elementos relevantes para la evaluación de casos de este dominio educativo).

Un Análisis Cognitivo de Tareas como el que se describe parte de la delimitación del dominio educativo, que debe implicar la identificación de las competencias deseables como desempeño en el dominio, la propuesta de indicadores de estas competencias, y posteriormente requiere de un trabajo fino de análisis en el cual se desglosan los objetivos (o competencias) del programa de estudio, y básicamente se identifican los niveles de complejidad: cognitiva, de tipo de conocimiento, del modelo mental a construir, etc., con un formato especial que se ha descrito por Castañeda (2006). Permite la especificación concreta de las habilidades que se evaluarán y se promoverán en la instrucción. Este análisis permite identificar cómo está conformado el conocimiento y, especialmente importante, cómo evaluar su estatus en términos del desempeño del aprendiz.

Con este esquema de evaluación se generan reactivos, actividades, simulaciones, sondeos, ejercicios y otras situaciones de evaluación interactiva en línea, sin perder de vista que cada elemento se puede ubicar dentro del esquema de análisis, y que al final se pueden realizar síntesis que arrojen una idea global del nivel de avance de cada aprendiz. La intención del uso de este esquema radica en que puede ser una alternativa de evaluación constante, que radiografe el proceso de aprendizaje.

En suma, el análisis permite identificar el nivel de complejidad de los componentes de las competencias. Evalúa conocimientos y habilidades a nivel micro, pero en el contexto de la competencia.

Castañeda (2004a) indica que la medición en sistemas orientados cognitivamente concibe los productos de aprendizaje como desarrollos graduales de habilidad cognitiva y plantea que el propósito de la medición es identificar el status actual del estudiante a partir de un modelo de desarrollo de la habilidad cognitiva. A partir del status y del tipo de errores detectados, será posible diseñar actividades de aprendizaje que fomenten el desarrollo cognitivo y los conocimientos pertinentes.

En el caso de construir instrumentos de evaluación con base en esta técnica, se sigue un procedimiento de validación y confiabilización, que consiste en lo siguiente:

1. el análisis inicial de competencias debe ser validado por expertos en el dominio, con algún instrumento y prueba de confiabilidad;
2. con un procedimiento similar se validan los niveles de demanda identificados para las tareas de las competencias;
3. con lo anterior se construyen los reactivos o ejercicios de evaluación, que una vez creados son sometidos también a la validación por jueces expertos, y se aplica una prueba estadística para determinar su grado de acuerdo;
4. cuando se cuenta con el instrumento, y antes de su utilización, se realiza un piloteo con una muestra de estudiantes, y a estos datos se aplican pruebas de calibración matemática, y
5. finalmente se depuran los instrumentos, que podrán convertirse a versiones Web e integrarse a ambientes de aprendizaje.

A partir de los elementos expuestos, en el presente trabajo se generó un conjunto de evaluaciones del desempeño de estudiantes inscritos en un curso en línea de psicología. El propósito del estudio fue poner a prueba una metodología para construir, validar y calibrar una batería de evaluaciones objetivas creadas con base en una técnica de Análisis Cognitivo de tareas.

2. MÉTODO

2.1. Participantes

Para el jueceo, los participantes fueron 10 profesores del área de Psicología Clínica de la carrera de psicología de la FES Iztacala, con una antigüedad promedio de 17 años, quienes han impartido al menos en diez ocasiones la asignatura de Psicología Clínica Teórica II, a la que corresponden los instrumentos desarrollados.

Para el piloteo y calibración de los instrumentos, los participantes fueron 70 alumnos de la licenciatura en psicología de la FESI-UNAM, inscritos en el sexto semestre de la carrera (43 mujeres y 27 hombres). Su edad promedio era de 22.3 años.

2.2. Situación y equipo

El jueceo se realizó en un cubículo de la Clínica Universitaria de Salud Iztacala, área en la que se concentra el cuerpo docente del área de Psicología Clínica. El cubículo tiene unas dimensiones de 3.5 por 2.5 metros, y cuenta con 1 computadora PC con procesador Pentium 4, 512 megabytes de memoria RAM, 120 Gigabytes de disco duro y acceso a Internet.

El piloteo de los instrumentos se realizó en una sala de cómputo ubicada en el Centro de Documentación de la Facultad de Estudios Superiores Iztacala. Las dimensiones de la sala son: 4 por 7 m² de superficie, paredes laterales de materiales prefabricados y un ventanal al frente. Esta sala cuenta con 40 computadoras PC con procesador Pentium 4, 256 megabytes de memoria RAM, 80 Gigabytes de disco duro y acceso a Internet.

2.3. Materiales

2.3.1. Construcción de los instrumentos

Para la construcción de los instrumentos, se utilizó la antología de la asignatura de Psicología Clínica Teórica II, desarrollada por un grupo de 12 profesores del área de Psicología Clínica de la carrera de Psicología de la FES Iztacala – UNAM.

La antología incluye dos volúmenes impresos: el primero es una guía didáctica, en la cual se incluyen: una introducción a la asignatura, así como introducciones a cada una de las tres unidades didácticas que la componen. En cada unidad se mencionan las referencias bibliográficas y se realiza una breve reseña de cada lectura. El segundo volumen es una recopilación impresa que integra reproducciones de todas las lecturas incluidas en el curso, agrupadas por unidad. Las unidades del curso son: 1) Introducción a la aproximación cognitiva conductual, con dos lecturas; 2) La evaluación cognitivo-conductual, con tres lecturas, y 3) El análisis funcional, con tres lecturas.

Se requirieron las siguientes herramientas para la creación de la versión en línea de los instrumentos de evaluación del curso:

Lenguaje HTML, para hacer las páginas Web que se requirieron para publicar el instrumento.

Base de datos MySQL, que permitió almacenar las respuestas a los reactivos por parte de cada uno de los participantes; lenguaje Java, con el objeto de programar los cuestionarios para el ambiente Web, y la

comunicación con la base de datos para almacenar las respuestas y posteriormente realizar reportes. Un servidor conectado a Internet en el cual hospedar el instrumento EDAOM en línea. Esto se hizo con el proveedor de almacenamiento para Internet www.performancehosting.net. Un software FTP (File Transfer Protocol) para transferir los archivos de la aplicación en el servidor del sitio Web: CuteFTP de GlobalScape.

FIGURA 1. MENÚ DE INGRESO A LAS EVALUACIONES DESDE EL META-TUTOR

Actividades de Samantha Arceo Rodriguez	
Psicología Clínica Teórica I	
Asignatura 2	
Evaluación 1	Evaluación inicial de opiniones acerca del estudio.
Evaluación 2	Evaluación inicial de conocimientos (no tiene ningún peso en la calificación).
Evaluación 3	Evaluación del tema: La aproximación Cognitiva Conductual.
Evaluación 4	Evaluación del tema: Evaluación y entrevista.
Evaluación 5	Evaluación del tema: Evaluación y Análisis funcional.
Tema 1	La aproximación cognitiva conductual
Tema 2	Evaluación y entrevista
Tema 3	Evaluación y análisis funcional

Meta-Tutor. Los instrumentos de evaluación en línea se incluyó como un componente dentro del ambiente de aprendizaje *Meta-Tutor* (Peñalosa y Castañeda, 2008b), que brindaba otros servicios instruccionales, pero al inicio presentaba un menú donde, entre otras opciones, permitía ingresar al EDAOM en línea cuando se programaba su aplicación.

FIGURA 2. MUESTRA DE UNO DE LOS INSTRUMENTOS DE EVALUACIÓN EN LÍNEA

Evaluación de Samantha Arceo Rodriguez

6 De acuerdo a la estructura de antecedentes del diagrama, elige la opción que engloba los factores a los que da lugar de manera directa la situación ilustrada en el siguiente texto:

"La psicoterapia americana estuvo muy influida por la psicología experimental. Empezando por John Watson e Ivan Pavlov, y continuando por Thorndike y Skinner, el centro de estudio se desplazó de la introspección y el estudio de la mente, hacia el desarrollo y examen de los principios del aprendizaje y el cambio de la conducta. En este proceso, la mente se redujo a una condición de epifenómeno (la "caja negra"), no porque no fuera importante o porque fuera inexistente, sino porque no podía observarse por otros individuos (Skinner, 1953). El conductismo radical subraya así dos principios explicativos: el condicionamiento clásico, basado en el aprendizaje por asociación, y el condicionamiento operante, basado en el aprendizaje por las consecuencias. La terapia conductual desarrolló intervenciones desde ambos principios y se convirtió en una fuerza importante en la psicología clínica americana durante 30 años. Sin embargo, diversos factores fueron responsables de su paulatina pérdida de importancia. En primer lugar, aunque las técnicas conductistas demostraban su eficacia, no lo hacían por encima de otras alternativas terapéuticas. Precisamente porque los psicólogos conductistas hicieron investigaciones empíricas, se desconcertaron cuando sus resultados mostraron diferencias no significativas con otras intervenciones. En segundo lugar, las técnicas conductistas no siempre funcionaban –contrariamente a las predicciones derivadas de las leyes del aprendizaje-. En otras palabras, [la "caja negra" entre los estímulos y las respuestas y sus consecuencias] parecía ser la responsable de mucha más variación de lo que podía pensarse! Por tanto, empezó la búsqueda de una explicación para esa variabilidad." (Dowd, 1997, pp. 26-27).

```

    graph TD
      A[Antecedentes de la Aproximación Cognitiva Conductual] --> B[Se dividen en:]
      B --> C[1 Histórico-Filosóficos]
      B --> D[2 Factores dentro de la psicología que propician su aparición]
      B --> E[3 Surgimiento de las terapias cognitivas]
    
```

1 Factores histórico-filosóficos
 2: desarrollo de factores dentro de la psicología que propician la aparición de la TCC
 3: surgimiento de las terapias cognitivas

Siguiente

La versión final de los instrumentos de evaluación de conocimiento académico aparecía en el *Meta-Tutor*, podía accederse a los mismos a través de un índice como el que se muestra en la figura 1, y un ejemplo de un reactivo de un instrumento aparece en la figura 2.

2.3.2. Jueceo de los instrumentos

Para el jueceo de los instrumentos, se requirió por un lado de los instrumentos ya desarrollados y disponibles en la dirección de Internet del *Meta-Tutor*. Se utilizó el formato de jueceo que se muestra en la tabla 1.

En el eje horizontal aparecen los números de los reactivos, y en el vertical las claves de cada uno de los jueces. Cada casilla es un espacio para la calificación del reactivo (cero o uno, dependiente del acuerdo o desacuerdo del juez con la aseveración que se evalúa).

TABLA1. FORMATO PARA JUECEO DE INSTRUMENTOS

	i1	i2	i3	i4	i5	i6	i7	i8	i9	i10	i11	i12	i13	i14	i15	i16	i17	i18	i19	i20	i21	i22	i23	i24	i25	i26	i27	
j1																												
j2																												
j3																												
j4																												
j5																												
j6																												
j7																												
j8																												
j9																												
j10																												

2.3.3. Calibración de reactivos

Para la calibración de los instrumentos, se utilizaron dos programas para PC: 1) ITEMAN (1993), un software para análisis de reactivos desarrollado de acuerdo con la teoría clásica de tests, de la casa *Assessment Systems Corporation*, y 2) RASCAL (1992), un software para calibración de reactivos de acuerdo con el modelo *Rasch* de un parámetro, de la Teoría de Respuesta al Ítem, de la misma casa de desarrollo de software.

2.4. Procedimiento

2.4.1. Construcción de los instrumentos

Los instrumentos que se construyeron fueron:

1. pre-postest, un cuestionario de 27 preguntas relacionadas con las tres unidades didácticas del curso de Psicología Clínica Teórica II (PCTII).
2. Evaluación de la unidad I, un cuestionario de 48 preguntas relacionadas con la primera unidad didáctica del curso de PCTII: la aproximación cognitivo conductual.
3. Evaluación de la *unidad 2*, un cuestionario de 40 preguntas relacionadas con la segunda unidad didáctica del curso de PCTII: la evaluación cognitivo conductual.

- Evaluación de la *unidad 2*, un cuestionario de 38 preguntas relacionadas con la tercera unidad didáctica del curso de PCTII: el análisis funcional.

Los instrumentos enumerados se construyeron siguiendo la lógica que se expone a continuación.

2.4.2. Prueba Pre-Post

Evaluaba en general el conocimiento del dominio, permitió conocer el nivel inicial de los participantes, y compararlo con su nivel final.

Pasos realizados en el desarrollo de la evaluación pre – post.

- Se realizó un análisis cognitivo de tareas como se describe en el apartado homónimo en este trabajo, a través del cual se identificaron los componentes de tarea a ser evaluados. Con ellos se construyó el diseño que dio el sustento teórico y puso los límites al universo de contenido que sería evaluado. Como productos derivados del diseño de observación, se establecieron los tres modelos mentales y los tres niveles de abstracción que representarían los contenidos cubiertos por la asignatura Psicología Clínica Teórica II de la carrera de psicología de la FES Iztacala, elegida como el área de interés.
- El diseño de observación representa los parámetros que rigieron la construcción de la evaluación. Éste se muestra en la tabla 2.

TABLA 2. ESPACIO DE CONTENIDOS DE LA EVALUACIÓN PRE - POST

Niveles de Complejidad cognitiva		Modelo mental		
		Conceptual	Estructural	Causal
	Solución de problemas	1	4	7
	Aplicación del conocimiento	2	5	8
	Comprensión	3	6	9

El dominio de conocimiento de la asignatura a impartir en el sistema de aprendizaje en línea se dividió en tres cortes, de acuerdo con lo expuesto en la sección que describe el ACT. Se generaron reactivos representativos del conocimiento para cada casilla en cada unidad. Son 3 unidades, por lo tanto para esta evaluación en total se generaron 27 reactivos.

2.4.3. Evaluaciones de las unidades 1, 2 y 3

La evaluación pre-post se desarrolló tomando en cuenta como ejes del diseño de observación tres modelos mentales y tres niveles de abstracción, todo lo anterior de tres unidades. En las evaluaciones de las unidades 1, 2 y 3, el diseño de observación tuvo una ligera diferencia: el universo del dominio se limitaba a los objetivos de cada una de las unidades, pero también tomando en cuenta los tres modelos mentales y los niveles de complejidad cognitiva.

Los tres instrumentos enumerados se generaron con la misma metodología, con base en la descripción del ACT, como se ilustra en la figura 1.

Como se ha mencionado, en esta figura se observan los gradientes de complejidad cognitiva y de modelos mentales, pero el tercer eje del modelo tridimensional se relaciona con la complejidad de contenidos de las unidades 1, 2 y 3 del curso. A continuación se describen los contenidos de estas unidades, así como los reactivos construidos para cada una.

Objetivos de las unidades:

Las unidades de aprendizaje tenían los siguientes objetivos, de acuerdo con la guía didáctica y antología del curso:

Unidad 1. La aproximación cognitivo conductual.

Objetivos. Al final del curso, el alumno:

- Conocerá los aspectos históricos de la Aproximación Cognitiva Conductual (ACC)
- Conocerá la definición y explicará los supuestos de la ACC
- Comprenderá las terapias y técnicas disponibles, sus usos y características

Se crearon dos reactivos por cada nivel de complejidad cognitiva, con dos niveles de dificultad en cada uno, y esto para cada uno de los modelos mentales y objetivos de la unidad excepto en los casos en los que no aplicara el cruce, de manera que se generaron 48 reactivos.

Unidad 2. La evaluación cognitivo - conductual

Las unidades de aprendizaje tenían los siguientes objetivos, de acuerdo con la guía didáctica y antología del curso:

- Caracterización de la Evaluación Conductual
- Instrumentos de evaluación de la TCC
- Conceptualización de la Entrevista conductual

Se crearon dos reactivos por cada nivel de complejidad cognitiva, con dos niveles de dificultad en cada uno, y esto para cada uno de los modelos mentales y objetivos de la unidad excepto en los casos en los que no aplicara el cruce, de manera que se crearon 40 reactivos.

Unidad 3. Evaluación y análisis funcional

Las unidades de aprendizaje tenían los siguientes objetivos, de acuerdo con la guía didáctica y antología del curso:

- Pasos en el desarrollo de la evaluación conductual
- Análisis funcional de casos

Se crearon dos reactivos por cada nivel de complejidad cognitiva, con dos niveles de dificultad en cada uno, y esto para cada uno de los modelos mentales y objetivos de la unidad. En este caso, sólo eran dos unidades, por lo tanto se generaron 36 reactivos.

2.4.4. Procedimiento de jueceo de los instrumentos

El procedimiento de jueceo fue el mismo en los cuatro casos, como sigue:

Se realizó la validación de los cuatro instrumentos de evaluación mediante la revisión por parte de un grupo de 10 jueces expertos en la asignatura, que son profesores con una experiencia de más de 10 años impartiendo la misma.

Se entregó el instrumento a los jueces, y se les dio un formato en el que debían indicar con un "1" si estaban de acuerdo con la validez de la pregunta, y un "0" en caso contrario.

Los jueces tuvieron acceso al diseño de observación, previamente elaborado, y en función de él realizaron sus observaciones acerca de la validez. Además de sus calificaciones de cada reactivo, todos los comentarios de los jueces se anotaron para efectos de optimización de la calidad de los instrumentos.

La prueba de confiabilidad de Kuder-Richardson-20, ilustrada en la Ecuación 2.1, que permite evaluar homogeneidad o consistencia interna en pruebas con datos dicotómicos, fue aplicada a los resultados del jueceo de los cuatro instrumentos, con el objeto de determinar el grado de acuerdo entre los jueces.

$$r_{kk} = \frac{k}{k-1} \left(\frac{s_x^2 - \sum p_i q_i}{s_x^2} \right) \quad (1)$$

En donde:

k= número de reactivos de la prueba

s_x^2 = varianza de los resultados de la prueba

p_i = proporción de jueces que aceptan un reactivo

$q_i = 1 - p_i$ = proporción de quienes no aceptan el reactivo

La prueba de Kuder-Richardson 20 entrega datos de la homogeneidad de la prueba; esto es, la consistencia interna de la misma (Brown, 1980); por lo tanto, serán útiles en este estudio para valorar si las opiniones de los jueces son consistentes.

2.4.5. Procedimiento de calibración de reactivos

La calibración es "un procedimiento psicométrico que permite poner en su punto de funcionamiento óptimo a los ítems de un instrumento, con base en un modelo de medición, de manera que al aplicarse a una población meta, aporten con adecuación, confiabilidad y utilidad, la información necesaria para tomar decisiones o medidas que, por ejemplo, conduzcan a la atención de un problema o a la formulación de un diagnóstico" (López, 2006a: 43).

El procedimiento general de calibración utilizado consistió en los siguientes pasos:

1. Al concluir el estudio, descargar del sitio Web del *Meta-Tutor* la información correspondiente a la lista de respuestas de 70 de los participantes ante cada uno de los instrumentos a calibrar.
2. Generar el archivo de base de datos para análisis, de acuerdo con los requerimientos de los programas para calibración descritos. Todos los programas requieren de una configuración de los datos para su lectura en archivos '.prn', con la distribución que se muestra en la tabla 3, donde las líneas: 1) primera línea, llamada de control, estipula: el número de reactivos, códigos para reactivos omitidos, no concluidos, así como la clave de usuarios; 2) en la segunda línea aparecen las respuestas correctas del instrumento; 3) en la tercera línea el número de opciones de respuesta, 4) en la cuarta línea se indica si se incluirá o no el reactivo en el análisis, y 5) de la quinta línea en adelante, aparecen todas las respuestas emitidas a cada reactivo por cada uno de los participantes.

TABLA 3. FORMATO DE ENTRADA DE DATOS PARA CALIBRACIÓN

27 0 N 0
 233323123232122231321131122
 333333333333333333333333333333
 YYYYYYYYYYYYYYYYYYYYYYYYYYYY
 311121233112132213332131123
 311332112233122322233322131
 212331133212131221223323133
 11112111213212323222322311
 112321233322112233223312223
 313331112232112223321333133
 311332212232111231122111221
 211312122112121233332112131

1. Ejecutar el programa en cuestión (dependiendo del análisis que se esté realizando), ITEMAN o RASCAL, indicando el nombre del archivo con los datos similar al descrito en la tabla 3, así como la ruta donde éste se encuentre en la computadora donde se ejecute el análisis.
2. El programa entrega un archivo con extensión '.out', que incluye la valoración de cada uno de los reactivos calibrados, de acuerdo con los parámetros que cada software maneja.

Los análisis realizados con los programas de calibración de reactivos enumerados anteriormente se describen a continuación.

ITEMAN y la Teoría Clásica de los Tests, o Teoría de Respuesta a la Prueba (TRP)

La Teoría Clásica de los Tests fue utilizada originalmente para desarrollar pruebas que determinaran la habilidad de las personas con respecto a ciertos reactivos incluidos en la prueba que pudieran satisfacer ciertas características y criterios predeterminados. En la TRP, las estadísticas de los reactivos dependen de las habilidades de los participantes; en otras palabras, los estudiantes con alta habilidad hacen que los reactivos se vean fáciles, y los alumnos con baja habilidad hacen que parezcan difíciles. Los índices de discriminabilidad se relacionan directamente con la homogeneidad del grupo de estudiantes evaluados.

ITEMAN Es un programa especializado en el análisis de reactivos de acuerdo con la Teoría de la Respuesta a la Prueba. López (2006b) describe a este tipo de programas indicando que ofrecen información útil acerca de los niveles de dificultad de cada reactivo, la discriminación entre desempeños altos y bajos, la distribución de respuestas en las opciones o distractores de cada reactivo, la correlación entre la respuesta al reactivo con respecto a las respuestas a toda la prueba, así como un sumario estadístico con los valores principales del instrumento aplicado, como el número de reactivos, examinados, media, desviación estándar, varianza, datos de la distribución, etc.

2.4.6. Programas de calibración según la Teoría de Respuesta al ITEM (TRI)

La TRI postula las siguientes seis ideas básicas: a) el desempeño de un evaluado en un reactivo de la prueba puede ser explicado por un conjunto de factores llamados rasgos, rasgos latentes o habilidades; b) los parámetros de los reactivos son independientes de la población de evaluados; en otras palabras, las estimaciones de la habilidad del evaluado no dependen de la prueba ni del grupo; c) los parámetros de habilidad son independientes de la elección de reactivos, los parámetros de reactivos y habilidades son independientes; d) la precisión de la estimación de habilidad puede determinarse en cada nivel de

habilidad; e) pueden seleccionarse reactivos para conformar una prueba con ciertas características, y f) la relación entre el desempeño de un participante en un reactivo y el conjunto de rasgos que subyacen al desempeño ante el reactivo pueden describirse mediante una función llamada función o curva característica del reactivo (Hambleton, Rogers y Swaminathan, 1991).

En la TRI, un reactivo puede caracterizarse por tres parámetros: a) el valor de discriminación, b) el nivel de dificultad, y c) el nivel de adivinación.

López (2006b) indica que estos modelos psicométricos permiten estimar el nivel de la habilidad de un examinando sobre el rasgo que está siendo medido, independientemente de la prueba específica que responda, asegurando que lo que se mide no dependa de con qué o a quién se mida.

El programa RASCAL se basa en el modelo Rasch de la TRI, cuyos supuestos son planteados por López (2006b):

1. Que la habilidad de los sustentantes puede determinarse de manera confiable, sin importar los reactivos específicos que se utilicen en una prueba en particular, por lo que se rompe la dependencia hacia el instrumento particular empleado para hacer la medición siempre y cuando los ítems hayan pasado por un proceso cuidadoso para su calibración. Este problema señala una de las mayores debilidades de la Teoría Clásica de Respuesta a la Prueba.
2. Que los reactivos calibrados tendrán también un comportamiento muy estable en sus valores de dificultad, independientemente de los sujetos particulares a los que se aplique el instrumento, en poblaciones que sean razonablemente similares.

El parámetro fundamental de análisis del programa RASCAL es la dificultad del reactivo, que implica el nivel de habilidad del participante. En términos generales, el rango de dificultad está entre -2 y 2, y los valores más allá de este rango representarán reactivos muy fáciles o muy difíciles. Se dice que esta escala permite identificar el ajuste a la prueba por parte de los evaluados, así como su habilidad.

El programa RASCAL permite también generar una curva característica de la prueba, en la que podrían identificarse los niveles de habilidad y la posibilidad de acertar.

El procedimiento de calibración incluyó el uso de los dos programas mencionados (ITEMAN y RASCAL) para el análisis de calibración de reactivos. Tomando en cuenta la propuesta de López (2006b) en el sentido de integrar en una sola tabla los indicadores de ambos modelos, en el presente estudio se integra información de calibración de los modelos de la TRP, así como de la TRI, con el objeto de tomar decisiones acerca de la constitución de los instrumentos.

TABLA 4. FORMATO DE CALIBRACIÓN CON INFORMACIÓN DE TRP Y TRI

Ítem	1	2	3	4	5	6	7	8	9	10
Dificultad	0.46	0.14	0.42	0.72	0.54	0.21	0.70	0.38	0.25	0.48
Discriminación	0.02	0.19	0.36	0.22	0.33	0.33	0.09	0.48	-0.06	0.36
Correlación	-0.03	0.16	0.25	0.29	0.30	0.40	0.01	0.38	0.06	0.27
Dífic. TRI	-0.03	1.63	0.14	-1.11	-0.32	1.14	-1.05	0.32	0.91	-0.09
Error Estándar	0.24	0.34	0.25	0.27	0.24	0.29	0.26	0.25	0.28	0.24

La tabla 4 es un ejemplo de cómo se reportan los resultados de las calibraciones. En la primera fila se incluye el número de reactivo, en las filas 2 a 4 la información derivada de los parámetros de la TRP y el software ITEMAN, y en las filas 5 y 6 la información de la TRI, obtenida con el software RASCAL. Con la

información de la tabla será posible identificar las propiedades de cada reactivo de una manera integrada, y en todo caso tomar las decisiones de la pertinencia o eliminación de reactivos en cada uno de los instrumentos a analizar.

3. RESULTADOS DEL JUECEO Y ANÁLISIS DE REACTIVOS

Toda evaluación enfrenta desafíos técnicos y tensiones que debe resolver. Para el caso de educación de adultos hay dificultades específicas relacionadas con una escasa cultura evaluativa, con nociones y prácticas erradas en evaluación, con la heterogeneidad y complejidad de la población que participa en las ofertas educativas, con la falta de formación docente y escasa reflexión pedagógica acerca de la evaluación.

3.1. Instrumento pre-post

3.1.1. Jueceo

La tabla 5 muestra las valoraciones de los jueces (j1.. j10) con respecto a los reactivos (i1 – i27) del instrumento pre-post:

TABLA 5. DATOS DE JUECEO DEL INSTRUMENTO PRE - POST

	i1	i2	i3	i4	i5	i6	i7	i8	i9	i10	i11	i12	i13	i14	i15	i16	i17	i18	i19	i20	i21	i22	i23	i24	i25	i26	i27
j9	0	1	1	1	1	1	1	0	1	1	1	1	0	1	1	0	1	1	1	0	1	1	1	1	1	1	1
j10	1	0	1	1	1	1	0	1	0	1	1	1	1	0	1	0	1	1	1	1	0	1	1	0	1	1	0
j3	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1
j7	1	1	0	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
j8	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
j1	1	1	1	1	1	1	1	1	0	1	1	0	1	1	1	1	1	0	1	1	1	0	1	1	1	1	1
j2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
j6	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
j5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
j4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

La tabla 6 muestra los resultados de las diferentes medidas utilizadas en la prueba Kuder-Richardson 20.

TABLA 6. VALORES DE LA PRUEBA KUDER-RICHARDSON 20 EN PRE-POSTEST

k	27
$\square pg_i$	1.74
\square	25
s_x^2	7
rkk	.83

Se observa que la prueba arrojó un coeficiente de confiabilidad de .83, que representa un valor aceptable de acuerdo entre jueces, e indica que no existen diferencias significativas entre sus opiniones respecto a la pertinencia de los reactivos.

TABLA 7. ANÁLISIS DE REACTIVOS DEL INSTRUMENTO PRETEST

Item	1	2	3	4	5	6	7	8	9	10	11	12	13
Dificultad	0.46	0.14	0.42	0.72	0.54	0.21	0.70	0.38	0.25	0.48	0.46	0.70	0.59
Discriminación	0.02	0.19	0.36	0.22	0.33	0.33	0.09	0.48	-0.06	0.36	0.19	0.21	0.16
Correlación	-0.03	0.16	0.25	0.29	0.30	0.40	0.01	0.38	0.06	0.27	0.20	0.31	0.11
<i>Dific. TRI</i>	<i>-0.03</i>	<i>1.63</i>	<i>0.14</i>	<i>-1.11</i>	<i>-0.32</i>	<i>1.14</i>	<i>-1.05</i>	<i>0.32</i>	<i>0.91</i>	<i>-0.09</i>	<i>-0.03</i>	<i>-1.05</i>	<i>-0.55</i>
Error Estándar	0.24	0.34	0.25	0.27	0.24	0.29	0.26	0.25	0.28	0.24	0.24	0.26	0.25

Item	14	15	16	17	18	19	20	21	22	23	24	25	26	27
Dificultad	0.42	0.46	0.79	0.34	0.30	0.56	0.42	0.28	0.37	0.39	0.34	0.87	0.45	0.28
Discriminación	-0.13	0.14	0.06	0.12	0.29	0.35	0.01	0.13	0.12	0.48	0.34	0.07	0.15	0.12
Correlación	0.05	0.15	0.08	0.04	0.24	0.34	-0.01	0.13	0.21	0.35	0.38	0.07	0.22	0.16
<i>Dific. TRI</i>	<i>0.14</i>	<i>-0.03</i>	<i>-1.49</i>	<i>0.50</i>	<i>0.70</i>	<i>-0.43</i>	<i>0.14</i>	<i>0.76</i>	<i>0.38</i>	<i>0.26</i>	<i>0.50</i>	<i>-2.10</i>	<i>0.02</i>	<i>0.76</i>
Error Estándar	0.25	0.24	0.29	0.26	0.26	0.24	0.25	0.27	0.25	0.25	0.26	0.35	0.24	0.27

Análisis de reactivos

Como se indicó en la sección de procedimiento, el análisis de reactivos del instrumento de evaluación pre-post fue realizado con ayuda del programa ITEMAN, así como el RASCAL, arrojando en resumen los datos que se observan en la tabla 7. Las filas en cursivas representan los valores del RASCAL, y las filas que no están en cursivas son valores del ITEMAN.

3.1.2. Ajustes al instrumento después del análisis

Debido a que esta aplicación de este instrumento corresponde a un conjunto de participantes con un estatus de principiantes, la dificultad identificada por ITEMAN es alta, con una media de .46; la discriminación es baja en promedio (.19), así como la correlación promedio de cada reactivo con el resto (.19). Por otro lado, la dificultad de acuerdo con la TRI está en un rango de -2.10 hasta 1.63, lo cual ubica a este rango como apto para identificar el ajuste de los evaluados.

Dado que un criterio para seleccionar reactivos es que la discriminación no sea negativa, se decidió eliminar los reactivos 9 y 14, que son negativos en este parámetro.

Por lo tanto, para este instrumento se dejan 25 reactivos en total.

3.2. Instrumento de evaluación de la Unidad I

Los resultados del jueceo del instrumento de evaluación de la Unidad 1 se muestran en la tabla 8, que se muestra abajo.

TABLA 8. DATOS DE JUECEO DEL INSTRUMENTO DE LA UNIDAD 1

	i1	i2	i3	i4	i5	i6	i7	i8	i9	i10	i11	i12	i13	i14	i15	i16	i17	i18	i19	i20	i21	i22	i23	i24
j1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
j2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
j3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0
j4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
j5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
j6	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
j7	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
j8	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
j9	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

j10	0	1	1	0	1	1	0	1	0	1	0	1	1	0	1	1	0	1	1	0	1	1	0	1
	i25	i26	i27	i28	i29	i30	i31	i32	i33	i34	i35	i36	i37	i38	i39	i40	i41	i42	i43	i44	i45	i46	i47	i48
j1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1
j2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
j3	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1
j4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
j5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
j6	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
j7	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
j8	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
j9	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
j10	0	1	1	0	1	1	0	1	1	1	1	1	1	1	0	1	1	1	0	1	1	1	0	0

La tabla 9 muestra los resultados de las diferentes medidas utilizadas en la prueba Kuder-Richardson 20.

TABLA 9. VALORES DE LA PRUEBA KUDER-RICHARDSON EN UNIDAD I

k	48
$\sum p_i q_i$	1.13
$\sum p_i^2$	42.40
sx ²	21.86
rkk	.97

Se observa que la prueba arrojó un coeficiente de confiabilidad de .97, que representa un grado importante de acuerdo entre jueces, e indica que no existen diferencias significativas entre sus opiniones respecto a la pertinencia de los reactivos.

3.2.1. Análisis de reactivos instrumento Unidad 1

Como se ha dicho antes, el instrumento de la Unidad 1 contaba con 48 reactivos, mismos que fueron sometidos al análisis ya descrito, donde se realizaron los procedimientos de calibración de acuerdo con la TRP (ITEMAN) y la TRI (RASCAL). Los resultados se muestran en la tabla 10.

En el análisis de acuerdo con el modelo de la TRP, realizado con el programa ITEMAN, La mayoría de los reactivos demostraron valores adecuados en función de los parámetros utilizados. Sin embargo, los reactivos 6, 27, 40 y 46 tuvieron una correlación negativa; los reactivos 27, 37, 40 y 46 tuvieron una discriminación negativa.

Por otro lado, al aplicar el análisis de acuerdo con el modelo de TRI, realizado con el programa RASCAL, los índices de discriminación fueron adecuados, excepto en los casos de los reactivos 6, 25, 32 y 46, con valores extremos de 4.01, -2.39, -2.39 y 2.32, respectivamente.

TABLA 10. ANÁLISIS DE REACTIVOS. INSTRUMENTO DE LA UNIDAD I

Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Dif.	0.27	0.85	0.52	0.44	0.88	0.05	0.64	0.79	0.56	0.79	0.50	0.80	0.79	0.74	0.58	0.68
Disc.	0.09	0.35	0.26	0.27	0.25	0.00	0.17	0.25	0.12	0.25	0.27	0.20	0.25	0.41	0.26	0.55
Corr.	0.22	0.39	0.20	0.26	0.35	-0.04	0.20	0.29	0.09	0.21	0.28	0.22	0.22	0.35	0.30	0.46
Dific. TRI	1.87	-1.03	0.75	1.08	-1.30	4.01	0.21	-0.60	0.55	-0.60	0.82	-0.69	-0.60	-0.33	0.49	-0.01
Error Est.	0.29	0.35	0.26	0.26	0.39	0.60	0.27	0.31	0.26	0.31	0.26	0.32	0.31	0.29	0.26	0.27

Item	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
Dif.	0.77	0.32	0.85	0.73	0.88	0.70	0.24	0.73	0.95	0.91	0.29	0.33	0.82	0.41	0.92	0.95
Disc.	0.45	0.28	0.15	0.26	0.15	0.45	0.24	0.50	0.15	0.05	-0.06	0.08	0.10	0.52	0.15	0.10
Corr.	0.49	0.30	0.33	0.34	0.15	0.52	0.18	0.42	0.33	0.03	-0.01	0.14	0.07	0.38	0.25	0.19
Dific. TRI	-0.50	1.64	-1.03	-0.24	-1.30	-0.08	2.04	-0.24	-2.39	-1.63	1.79	1.56	-0.80	1.21	-1.84	-2.39
Error Est.	0.30	0.27	0.35	0.29	0.39	0.28	0.30	0.29	0.60	0.44	0.28	0.27	0.33	0.26	0.47	0.60

Item	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48
Dif.	0.83	0.38	0.76	0.95	0.77	0.68	0.39	0.58	0.89	0.58	0.74	0.88	0.55	0.20	0.82	0.42
Disc.	0.25	0.41	0.01	0.15	-0.04	0.41	0.27	-0.12	0.10	0.36	0.40	0.10	0.17	-0.01	0.30	0.17
Corr.	0.40	0.26	0.00	0.20	0.10	0.39	0.26	-0.01	0.24	0.28	0.40	0.16	0.19	-0.06	0.27	0.12
Dific. TRI	-0.91	1.35	-0.41	-2.39	-0.50	-0.01	1.28	0.49	-1.46	0.49	-0.33	-1.30	0.62	2.32	-0.80	1.15
Error Est.	0.34	0.26	0.30	0.60	0.30	0.27	0.26	0.26	0.41	0.26	0.29	0.39	0.26	0.32	0.33	0.26

3.2.2. Ajustes al instrumento de la Unidad 1 después del análisis

Dados los datos indicados anteriormente, se decidió eliminar los siguientes reactivos: 6, 25, 27, 32, 36, 37, 40 y 46.

Por lo tanto, este instrumento se conformó de 40 reactivos.

3.3. Instrumento de evaluación de la Unidad 2

Los resultados del jueceo del instrumento de evaluación de la Unidad 2 se muestran en la tabla 11.

TABLA 11. DATOS DE JUECEO DEL INSTRUMENTO DE LA UNIDAD 2

	i1	i2	i3	i4	i5	i6	i7	i8	i9	i10	i11	i12	i13	i14	i15	i16	i17	i18	i19	i20
j1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
j2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
j3	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1
j4	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1
j5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
j6	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
j7	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
j8	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1
j9	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
j10	1	1	1	1	1	1	0	1	1	0	1	1	1	1	0	1	1	1	0	1

	i21	i22	i23	i24	i25	i26	i27	i28	i29	i30	i31	i32	i33	i34	i35	i36	i37	i38	i39	i40
j1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	0
j2	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1
j3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
j4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
j5	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1
j6	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
j7	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1
j8	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
j9	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	1	0	1	1	1
j10	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	0	1	1

La tabla 12 muestra los resultados de las diferentes medidas utilizadas en la prueba Kuder-Richardson 20.

TABLA 12. VALORES DE LA PRUEBA KUDER-RICHARDSON EN UNIDAD 2

k	40
$\sum p_i q_i$	0.73
$\sum x_i^2$	34.8
sx ²	14.26
rkk	0.973

Se observa que la prueba arrojó un coeficiente de confiabilidad de .97, que representa un grado importante de acuerdo entre jueces, e indica que no existen diferencias significativas entre sus opiniones respecto a la pertinencia de los reactivos.

3.3.1. Análisis de reactivos del instrumento Unidad 2

El instrumento de la *Unidad 2* contaba con 40 reactivos, mismos que fueron sometidos al análisis de calibración de acuerdo con la TRP (ITEMAN) y la TRI (RASCAL). Los resultados se muestran en la tabla 13.

TABLA 13. ANÁLISIS DE REACTIVOS DEL INSTRUMENTO DE LA UNIDAD 2

Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Dif.	0.85	0.80	0.24	0.48	0.98	0.79	0.76	0.58	0.89	0.79	0.70	0.56	0.50	0.21
Disc.	0.17	0.41	-0.05	0.52	0.05	0.35	0.30	0.72	0.23	0.17	0.28	0.16	0.33	-0.30
Corr.	0.16	0.44	0.12	0.42	0.26	0.22	0.28	0.57	0.37	0.22	0.24	0.12	0.26	-0.22
Dific. TRI	-1.18	-0.85	1.89	0.72	-3.68	-0.75	-0.57	0.33	-1.61	-0.75	-0.24	0.39	0.66	2.08
Error Est.	0.35	0.32	0.30	0.26	1.01	0.31	0.30	0.26	0.41	0.31	0.28	0.26	0.26	0.31

Item	15	16	17	18	19	20	21	22	23	24	25	26	27	28
Dif.	0.62	0.62	0.77	1.00	0.41	0.59	0.62	0.39	0.55	0.65	0.89	0.44	0.52	0.42
Disc.	0.44	0.52	0.16	0.00	0.32	0.28	0.03	0.44	0.11	0.32	0.18	0.48	0.42	0.66
Corr.	0.30	0.42	0.11	0.00	0.31	0.23	0.10	0.39	0.16	0.33	0.24	0.41	0.34	0.57
Dific. TRI	0.12	0.12	-0.66	borrar	1.06	0.26	0.12	1.12	0.46	-0.02	-1.61	0.92	0.59	0.99
Error Est.	0.26	0.26	0.30		0.26	0.26	0.26	0.26	0.26	0.27	0.41	0.26	0.26	0.26

Item	29	30	31	32	33	34	35	36	37	38	39	40
Dif.	0.35	0.50	0.77	0.47	0.91	0.62	0.67	0.67	0.76	0.76	0.65	0.42
Disc.	0.20	0.31	0.22	0.18	0.17	0.29	0.50	-0.04	0.15	0.21	-0.10	0.40
Corr.	0.14	0.20	0.14	0.20	0.25	0.25	0.45	0.01	0.17	0.21	0.01	0.42
Dific. TRI	1.34	0.66	-0.66	0.79	-1.78	0.12	-0.09	-0.09	-0.57	-0.57	-0.02	0.99
Error Est.	0.27	0.26	0.30	0.26	0.43	0.26	0.27	0.27	0.30	0.30	0.27	0.26

En el análisis de acuerdo con el modelo de la TRP, realizado con el programa ITEMAN, el reactivo 14 tuvo una correlación negativa con el resto de los reactivos; los reactivos 3, 14, 36 y 39 tuvieron una discriminación negativa. El reactivo 18 tuvo una dificultad de 1, ya que todos los participantes de la muestra lo respondieron correctamente.

Por otro lado, al aplicar el análisis de acuerdo con el modelo de TRI, realizado con el programa RASCAL, los índices de discriminación fueron adecuados, excepto en el reactivo 14, con un valor extremo de 2.08, lo que lo hacía demasiado difícil.

3.3.2. Ajustes al instrumento de la Unidad 1 después del análisis

Dados los datos indicados anteriormente, se decidió eliminar los siguientes reactivos: 3, 5, 14, 18, 36 y 39.

Por lo tanto, este instrumento se conformó de 34 reactivos.

3.4. Instrumento de evaluación de la Unidad 3

Los resultados del jueceo del instrumento de evaluación de la Unidad 3 se muestran en la tabla 14.

TABLA 14. DATOS DE JUECEO DEL INSTRUMENTO DE LA UNIDAD 3

	i1	i2	i3	i4	i5	i6	i7	i8	i9	i10	i11	i12	i13	i14	i15	i16	i17	i18	i19
j1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
j2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
j3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
j4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
j5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
j6	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
j7	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
j8	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
j9	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
j10	1	1	1	1	0	1	1	1	1	0	1	1	1	1	1	1	0	1	1

	i20	i21	i22	i23	i24	i25	i26	i27	i28	i29	i30	i31	i32	i33	i34	i35	i36
j1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1
j2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
j3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
j4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
j5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
j6	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
j7	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
j8	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1
j9	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1
j10	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1

La tabla 15 muestra los resultados de las diferentes medidas utilizadas en la prueba Kuder-Richardson.

TABLA 15. VALORES DE LA PRUEBA KUDER-RICHARDSON 20 EN UNIDAD 2

k	36
\square piqi	0.45
\square	32.1
sx2	11.65
rkk	0.989

Se observa que la prueba arrojó un coeficiente de confiabilidad de .989, que representa un grado importante de acuerdo entre jueces, e indica que no existen diferencias significativas entre sus opiniones respecto a la pertinencia de los reactivos.

3.4.1. Análisis de reactivos del instrumento Unidad 3

El instrumento de la Unidad 3 contaba con 38 reactivos, mismos que fueron sometidos al análisis de calibración de acuerdo con la TRP (ITEMAN) y la TRI (RASCAL). Los resultados se muestran en la tabla 16.

En el análisis de acuerdo con el modelo de la TRP, realizado con el programa ITEMAN, los reactivos 14 y 36 tuvieron una correlación negativa con el resto de los reactivos (-0.19 y -0.06, respectivamente); los reactivos 14, 21 y 36 tuvieron una discriminación negativa (-0.29, -0.03 y -0.12, respectivamente). El reactivo 18 tuvo una dificultad de 1, ya que todos los participantes de la muestra lo respondieron correctamente.

TABLA 16. ANÁLISIS DE REACTIVOS DEL INSTRUMENTO DE LA UNIDAD 3

Item	1	2	3	4	5	6	7	8	9	10	11	12	13
Dif.	0.85	0.80	0.24	0.48	0.98	0.79	0.76	0.58	0.89	0.79	0.70	0.56	0.50
Disc.	0.15	0.40	0.00	0.59	0.04	0.29	0.39	0.65	0.24	0.19	0.22	0.22	0.39
Corr.	0.18	0.43	0.10	0.42	0.23	0.25	0.31	0.57	0.41	0.27	0.25	0.13	0.30
Dific. TRI	-1.19	-0.86	1.90	0.72	-3.69	-0.76	-0.58	0.32	-1.62	-0.76	-0.25	0.39	0.65
Error Est.	0.35	0.32	0.30	0.26	1.01	0.31	0.30	0.26	0.41	0.31	0.28	0.26	0.26

Item	14	15	16	17	18	19	20	21	22	23	24	25
Dif.	0.21	0.62	0.62	0.77	1.00	0.41	0.59	0.62	0.39	0.55	0.65	0.89
Disc.	-0.29	0.49	0.44	0.08	0.00	0.36	0.28	-0.03	0.46	0.23	0.34	0.12
Corr.	-0.19	0.32	0.40	0.12		0.31	0.24	0.10	0.40	0.19	0.33	0.18
Dific. TRI	2.09	0.11	0.11	-0.67	borr	1.06	0.25	0.11	1.13	0.45	-0.03	-1.62
Error Est.	0.31	0.26	0.26	0.30		0.26	0.26	0.26	0.26	0.26	0.27	0.41

Item	26	27	28	29	30	31	32	33	34	35	36
Dif.	0.44	0.52	0.42	0.35	0.50	0.77	0.47	0.91	0.62	0.67	0.67
Disc.	0.52	0.42	0.59	0.25	0.31	0.21	0.14	0.11	0.31	0.44	-0.12
Corr.	0.38	0.38	0.57	0.17	0.25	0.16	0.17	0.20	0.23	0.43	-0.06
Dific. TRI	0.92	0.59	0.99	1.34	0.65	-0.67	0.79	-1.79	0.11	-0.10	-0.10
Error Est.	0.26	0.26	0.26	0.27	0.26	0.30	0.26	0.43	0.26	0.27	0.27

Por otro lado, al aplicar el análisis de acuerdo con el modelo de TRI, realizado con el programa RASCAL, los índices de discriminación fueron adecuados, excepto en el reactivo 5, con un valor extremo de -3.69, que lo hacía demasiado fácil.

3.4.2. Ajustes al instrumento de la Unidad 3 después del análisis

Dados los datos indicados anteriormente, se decidió eliminar los siguientes reactivos: 3, 5, 14, 18, 21 y 36. Por lo tanto, este instrumento se conformó de 31 reactivos.

3.5. Resumen de resultados de todos los instrumentos analizados

Como se indica en los análisis presentados antes, los datos de jueceo fueron favorables. Un resumen de los valores obtenidos en los análisis de confiabilidad de los jueceos se encuentra en la tabla 17 se muestran los resultados de la prueba Kuder-Richardson para todos los instrumentos evaluados. Estos datos fueron satisfactorios.

TABLA 17. TABLA RESUMEN DE CONFIABILIDAD EN JUECEO

Instrumento	Pre-post	Unidad 1	Unidad 2	Unidad 3
kkrr	.83	.97	.97	.99

TABLA 18. REACTIVOS ACEPTADOS DESPUÉS DEL ANÁLISIS DE CALIBRACIÓN

Instrumento	Pre-post	Unidad 1	Unidad 2	Unidad 3
Núm. reactivos	25	40	34	31
Eliminados	2	8	6	5
Reactivos que integran el instrumento:	1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13 15, 16, 17, 18, 19 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30	1, 2, 3, 4, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 26, 28, 29, 30, 31, 33, 34, 35, 38, 39, 41, 42, 43, 44, 45, 47, 48	1, 2, 4, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, 17, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 37, 38, 40	1, 2, 3, 4, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, 17, 19, 20, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35

Por otro lado, los resultados de las calibraciones se incluyen en la tabla 18, en la que se muestran los reactivos seleccionados en cada instrumento.

Con estos análisis se tuvieron instrumentos que cumplieron con criterios metodológicos que permiten inferir su validez y confiabilidad, con el objeto de utilizarlos en el proceso instruccional del curso descrito.

4. DISCUSIÓN

La metodología de construcción de instrumentos a utilizar en el sistema instruccional en línea se basó en el análisis cognitivo de tareas, e incorporó como dimensiones de análisis: la complejidad cognitiva de los procesos a evaluar, los modelos mentales a construir por parte de los estudiantes, y las objetivos de las unidades de aprendizaje de los programas del curso que se seleccionó para este trabajo.

Los datos de jueceo de los instrumentos confirman su validez y la pertinencia de los reactivos en términos de su apego a los contenidos y objetivos del programa. Por otro lado, los análisis de calibración permitieron, desde las teorías psicométricas clásica y de respuesta al ítem, seleccionar los reactivos que permitían identificar de una forma más confiable los conocimientos de los estudiantes y descartar reactivos dada su falta de discriminación o correlación, o bien por tener un nivel de dificultad extremo.

Las implicaciones de estos resultados se relacionan con la importancia de contar con una metodología para la construcción de instrumentos de evaluación objetiva que permita realizar diagnósticos de los niveles de desempeño de los estudiantes en contextos académicos como el que se describe en el presente trabajo. Especialmente en el contexto de la educación mediante Internet, en la que es preciso construir ambientes que confieran a la evaluación la función doble: de apoyo al aprendizaje, y al mismo tiempo de diagnóstico del desempeño. El presente trabajo contribuye en la medida en que presenta esquemas de evaluación útiles para valorar en la investigación los impactos de una diversidad de componentes incluidos en los ambientes de aprendizaje en línea, lo cual es importante ante la escasez de estudios que realmente permiten valorar los elementos que determinan el desempeño en estos contextos (Sadik, 2001), y ante la consideración de que la evaluación es un "punto débil" de la educación virtual (Barberá, 2006).

El modelo que se presenta adapta una metodología probada en el área de la evaluación del aprendizaje complejo en la educación superior presencial, para incorporarla a los ambientes virtuales como un mecanismo que puede dar estructura a las evaluaciones en un ambiente de aprendizaje en línea, partiendo de la "disección" de las competencias. De esta forma, las competencias y sus componentes, que representan las metas de aprendizaje de los programas educativos, son el eje de la evaluación.

Una vez analizados los contenidos e identificadas las competencias del dominio educativo, el Análisis Cognitivo de Tareas conduce a un potente esquema de evaluación que permite diagnosticar la estructuración del conocimiento de los estudiantes a lo largo de cursos en línea, tomando en cuenta dimensiones centrales de sus procesos cognitivos. La creación del modelo del dominio de conocimiento de los cursos conduce a la construcción de instrumentos de evaluación válidos y confiables.

Modelar estas estructuras puede ser útil en la investigación del aprendizaje en línea, ya que ayuda a determinar la influencia de variables. Investigaciones de este tipo se requieren en el campo de la educación a distancia (Jegede , 1999), y en línea (Quesada, 2006; Barberá, 2006). La ausencia de

investigación acerca de la evaluación del aprendizaje es evidente, y propuestas de este tipo podrían ayudar a generar conocimiento y métodos de estudio en el área.

En suma, la presente propuesta podría aplicarse para realizar investigación acerca de diversos métodos de evaluación y fomento del aprendizaje en línea, gracias a la seguridad que ofrece en el sentido de delimitar el dominio educativo bajo estudio mediante una taxonomía válida y sensible.

La educación en línea es considerada una opción flexible, efectiva y en tanto viable para cubrir la alta demanda educativa de nuestro país, pero si no generamos investigación que explique los efectos de nuestras intervenciones de evaluación y fomento del aprendizaje en estos contextos, en este pragmatismo podríamos dirigirnos hacia un escenario caracterizado por malos resultados, que es lo que ya no queremos en la educación mexicana.

REFERENCIAS BIBLIOGRÁFICAS

- Astolfi, J.P. (2003). *Aprender en la escuela*. Santiago de Chile: LOM.
- Anderson, T. (2004). Toward a theory of online learning. En T. Anderson y F. Elloumi (Eds), *Theory and practice of online learning* (pp. 33-58). Athabasca: Athabasca University.
- Barberá, E. (2006). Aplicaciones de la tecnología a la e-Evaluación. *Revista de Educación a Distancia*, 5, Consultado el 27 de julio de 2008 en: <http://www.um.es.ead/red/M6>.
- Bransford, J., Brown, A. L., y Cockings, R. R. (2004). *How people learn: Brain, mind, experience, and school. Expanded edition*. Washington, D.C.: National Academy Press.
- Brown, F. (1980). *Principios de la medición en Psicología y Educación*. Bogotá: Manual Moderno.
- Castañeda, S. (1998a). *Evaluación y fomento del desarrollo intelectual en la enseñanza de ciencias, artes y técnicas: Perspectiva internacional en el umbral del siglo XXI*. México: Editorial Miguel Angel Porrúa - UNAM.
- Castañeda, S. (1998b). Evaluación de resultados de aprendizaje en escenarios educativos. *Revista Sonorense de Psicología*, 12, pp. 57-67.
- Castañeda, S. (2002). A cognitive model for learning outcomes assessment. *International Journal of Continuing Engineering Education and Life-long Learning*, 12, pp. 94-106.
- Castañeda, S. (2004a). Educación, aprendizaje y cognición. En S. Castañeda, S. (Ed). *Educación, aprendizaje y cognición: teoría en la práctica* (pp. 49-74). México: Manual Moderno.
- Castañeda, S. (2004b). Enseñanza estratégica: guía abreviada para el docente y el tutor En S. Castañeda, S. (Ed), *Educación, aprendizaje y cognición: teoría en la práctica* (pp. 349-421). México: Manual Moderno.
- Castañeda, S. (2006). Evaluación del aprendizaje en educación superior. En: S Castañeda (coord). *Evaluación del aprendizaje en el nivel universitario: elaboración de exámenes y reactivos objetivos* (pp. 3-27). México: UNAM.

- Chipman, S. F., Schraagen, J. M., Shalin, V. L. (2000). Introduction to Cognitive Task Analysis. En J. M. Schraagen, S. F. Chipman y V. L. Shalin (eds.), *Cognitive Task Analysis* (pp. 3-23). New York: Lawrence Erlbaum Associates.
- Hambleton, R.K., Rogers, H.J. y Swaminathan, H. (1991). *Fundamentals of item response theory*. Londres: Sage Publications.
- ITEMAN (1993). *Conventional Item Analysis Program*. (Versión 3.5). St. Paul, MN: Assessment Systems Corp.
- López, M. (2006a). Preparación y calibración de reactivos: medición de y para el aprendizaje. En S. Castañeda (Coord.) *Evaluación del Aprendizaje en el Nivel Universitario: Elaboración de exámenes y reactivos objetivos* (pp. 29-56). México: UNAM.
- López, O. M. (2006b). Análisis de reactivos. En S. Castañeda (Coord.), *Evaluación del Aprendizaje en el Nivel Universitario: Elaboración de exámenes y reactivos objetivos* (pp. 215-245) México: UNAM.
- Peñalosa, E. y Castañeda, S. (2008a). Generación de conocimiento en la educación en línea: un modelo para el fomento de aprendizaje activo y autorregulado. *Revista Mexicana de Investigación Educativa*, 13(36), 229-260.
- Peñalosa, E. y Castañeda, S. (2008b). Meta-Tutor: an online environment for knowledge construction and self-regulated learning in clinical psychology teaching. *International Journal of Continuing Education, Engineering and Lifelong Learning*, 18(3), pp. 283-297.
- Quesada, R. (2006). Evaluación del aprendizaje en la educación a distancia en línea. *Revista de Educación a Distancia*, 5, Consultado el 27 de julio de 2008 en <http://www.um.es/ead/red/M6>.
- RASCAL (1992). *Rasch Analysis Program* (Versión 3.5). St. Paul, MN: Assessment Systems Corp.
- Renkl, A., Mandl, H. y Gruber, H. (1996). Inert knowledge: Analyses and remedies. *Educational Psychologist*, 3, pp. 115-121.
- Sadik, A. (2001). *Directions for future research in on line distance education*. Documento de disertación doctoral, Facultad de Educación de la Universidad South Valley, Qena 11183, Egipto.
- Schwartz, D.L., Brophy, S., Lin, X. y Bransford, J.D. (1999). Software for managing complex learning: examples from an educational psychology course. *Educational Technology Research and Development*, 47, pp. 39-60.
- Van Merriënboer, J.G., Clark, R. y deCrook, M.B.M (2002). Blueprints for complex learning: the 4 C/ID – Model. *Education, Training and Development*, 50, pp. 39-64.