



Universidad Autónoma de Baja California
Instituto de Investigación y Desarrollo Educativo

**“Desarrollo de un módulo de capacitación y asesoría
para apoyar el análisis, reporte e interpretación de
resultados derivados de la aplicación de pruebas
criteriales de gran escala”.**

T E S I S

Que para obtener el grado de
MAESTRA EN CIENCIAS EDUCATIVAS

Presenta

Citlalli Sánchez Álvarez

Ensenada, B. C., septiembre de 2007.



Universidad Autónoma de Baja California
Instituto de Investigación y Desarrollo Educativo
Maestría en Ciencias Educativas



“Desarrollo de un módulo de capacitación y asesoría para apoyar el análisis, reporte e interpretación de resultados derivados de la aplicación de pruebas criteriosales de gran escala”.

T E S I S

Que para obtener el grado de
MAESTRA EN CIENCIAS EDUCATIVAS

Presenta
Citlalli Sánchez Alvarez

APROBADA POR:

M. C. Luis Ángel Contreras Niño
Director de Tesis

Dr. Joaquín Caso Niebla
Sinodal

M. C. Virginia Velasco Ariza
Sinodal

M. C. Ana Isabel Miramontes Bush
Sinodal

Ensenada, B. C., septiembre de 2007.

Índice de contenidos

Página

Capítulo I.- INTRODUCCIÓN.	1
1.1. Antecedentes.	1
1.2. Planteamiento del problema.	8
1.3. Objetivos.	11
1.4. Justificación.	12
1.5. Supuestos.	13
Capítulo II.- ANTECEDENTES TEÓRICO-CONCEPTUALES Y CONTEXTO PSICOMÉTRICO DE LA EVALUACIÓN DEL APRENDIZAJE.	15
2.1. Aspectos conceptuales generales sobre la evaluación psicométrica.	15
2.2. Tipo de pruebas que evalúan el aprendizaje.	16
2.2.1. Similitudes y diferencias entre pruebas normativas y pruebas criterioles.	20
2.2.2. Determinación del tipo de prueba que se utilizará.	22
2.2.3. Uso de las pruebas criterioles y de sus resultados en la elaboración de informes de resultados.	23
2.3. Normatividad internacional y nacional para la elaboración, evaluación y uso de pruebas.	25
2.4. Modelos psicométricos para evaluar el aprendizaje.	29
2.5. Establecimiento de estándares de ejecución y puntos de corte.	35
Capítulo III.- ESQUEMA ANALÍTICO – INFORMÁTICO PARA ELABORAR INFORMES DE RESULTADOS.	43
3.1. Análisis curricular.	43
3.2. Base de datos del examen.	45
3.3. Variables de contexto.	47
3.4. Análisis de la ejecución de los examinandos.	51
3.5. Destinatarios de informes de resultados.	52
3.5.1. Análisis, reporte e interpretación de resultados en forma desagregada.	60
3.6. Esquema informático.	72
Capítulo IV.- PROCEDIMIENTOS PARA CAPACITAR AL CARIRE EN EL ANÁLISIS DE RESULTADOS Y LA ELABORACIÓN DE INFORMES.	73
4.1. Participantes.	73
4.2. Instrumentos.	74
4.3. Procedimientos.	75
Capítulo V.- CONCLUSIONES Y RECOMENDACIONES PARA LA OPERACIÓN Y EVALUACIÓN DEL MÓDULO DE CAPACITACIÓN Y ASESORÍA.	78
BIBLIOGRAFÍA.	82
ANEXOS.	88

Índice de tablas

Página

Tabla 1.	Diferencias entre pruebas referidas a una norma y referidas a un criterio.	20
Tabla 2.	Modelos continuos más comunes para establecer puntos de corte.	42
Tabla 3.	Destinatarios de los informes de resultados del examen.	61

Índice de figuras

Página

Figura 1. Esquema analítico - informático de los resultados de la aplicación a gran escala de un examen criterial. 44

Figura 2. Ejemplo de gráfica de cajas y bigotes. 67

Índice de anexos

Página

Anexo 1.	Manual del participante del Módulo 5: Análisis, Reporte e Interpretación de Resultados de Pruebas Criteriales de Gran Escala.	90
Anexo 2.	Formato para elaborar el informe de resultados del examen.	174
Anexo 3.	Ejemplos para ilustrar la aplicación de los procedimientos y criterios para el análisis de datos y la elaboración de informes de resultados.	178
Anexo 4.	Cuestionario para la evaluación de la operación del módulo.	189
Anexo 5.	Justificación del cuestionario de evaluación de la operación del módulo.	193
Anexo 6.	Formato para cotejar los componentes del análisis de la ejecución de los examinandos.	201
Anexo 7.	Formato para cotejar los componentes del informe de resultados.	204

Agradecimientos

A mi madre Eva Luz con el más delicado y profundo amor, por tantos años de ser el ejemplo a seguir; por su entrañable comprensión, incansable apoyo y tierno amor; por ser un modelo de integridad, determinación y fuerza durante toda mi vida; y especialmente, por tenerme siempre en su corazón.

A mis hermanas Eva Luz y Ariadna, por ser siempre mi hogar; y a mis sobrinos por ser quienes aprenderán de la sabiduría de varias generaciones.

A Luis Ángel Contreras Niño, mi querido Director de tesis, quien con su inagotable paciencia, sabios consejos y cálida amistad guió mi formación no solamente académica, sino personal; y a quien admiro por su brillante complejidad, ingeniosa sencillez, y gran calidad humana.

A mis maestros por compartir sus conocimientos y esperar siempre lo mejor de sus alumnos.

Capítulo I.- INTRODUCCIÓN

1.1. Antecedentes

El interés por realizar este estudio partió de la labor que he realizado como colaboradora de un equipo de investigación que ha desarrollado desde 1997 un modelo para elaborar pruebas criterios de gran escala alineadas con el currículum, mismo que ha sido empleado desde septiembre de 2003 para capacitar y asesorar a personal académico de varias instituciones educativas de nivel superior y medio superior, entre las cuales figuran el Colegio de Bachilleres de Baja California (COBACH), el Colegio de Estudios Científicos y Tecnológicos de Baja California (CECyTE) y la Universidad Autónoma de Baja California (UABC), para elaborar exámenes sumarios, de seguimiento y colegiados que aporten elementos para la certificación y el monitoreo del aprendizaje de sus estudiantes. El interés personal surgió también del hecho de que dicho modelo no se encuentra acabado y aún requiere de importantes desarrollos que lo fortalezcan, mismos que constituyen áreas de oportunidad.

En el año 2001 se publicó el programa nacional de educación (PNE) 2001-2006, el cual proponía una reforma al sistema educativo mexicano, en donde se planteaban objetivos particulares, líneas de acción y metas. Uno de dichos objetivos fue llevar a cabo un fortalecimiento del Sistema Educativo Nacional, mediante la consolidación del sistema de evaluación. Para ello, propuso crear el Instituto

Nacional para la Evaluación de la Educación, instituir el organismo de evaluación para la educación media superior, y establecer criterios relativos al uso y la difusión de la evaluación (Secretaría de Educación Pública, 2001).

El PNE destacaba la preocupación reciente que ha surgido en torno al perfeccionamiento de la medición de los indicadores de eficiencia y, especialmente, la medición del aprovechamiento escolar, y que aún prevalece “el uso insuficiente de los instrumentos y sistemas de evaluación, y la inexistencia de mecanismos claros y sistemáticos de rendición de cuentas –relativos al sistema y su organización, al aprendizaje de los alumnos, al impacto de los programas y el desempeño del personal, entre otros– los cuales contribuyen a desalentar el funcionamiento eficiente del sistema” (Secretaría de Educación Pública, 2001 p. 121).

En respuesta a lo anterior, instituciones educativas de nivel medio superior y superior desarrollaron estrategias para atender las necesidades educativas nacionales, que derivaran en acciones para abordar y dar solución a problemáticas como la planteada anteriormente. En el estado de Baja California, las principales acciones institucionales emprendidas para mejorar la evaluación del aprendizaje a gran escala fueron realizadas principalmente por el Colegio de Bachilleres de Baja California (COBACH), el Colegio de Estudios Científicos y Tecnológicos de Baja California (CECyTE) y la Universidad Autónoma de Baja California (UABC).

En junio de 2003, la UABC publicó el Plan de Desarrollo Institucional (PDI) 2003-2006 el cual alude, dentro de su iniciativa denominada: Fortalecimiento de los

procesos y mecanismos institucionales de seguimiento y evaluación, a los procesos y mecanismos de seguimiento y evaluación del funcionamiento de los procesos académicos y administrativos. El propósito fundamental de esta iniciativa fue promover la creación de un sistema de indicadores y de un esquema de evaluación que permitieran dar seguimiento al desarrollo de diferentes procesos institucionales, en particular los académicos, fomentando un uso cotidiano de la información generada y que la toma de decisiones se beneficie de ello (Universidad Autónoma de Baja California, 2003).

Dentro de las iniciativas específicas del PDI, se encontraba la relacionada con la evaluación institucional de la calidad del aprendizaje, en donde se menciona que para lograr una educación de buena calidad se requiere conocer de forma sistemática los logros alcanzados en el proceso de aprendizaje por los alumnos, y propuso diseñar un programa de evaluación del aprendizaje que proporcionara la información necesaria para mejorar y asegurar la calidad académica de la institución y rendir cuentas a la sociedad en general. Dentro de los objetivos propuestos para el programa se encontraba el contar con información sistematizada sobre la calidad de los aprendizajes logrados, para mejorar el proceso educativo; y para ello propuso:

- Desarrollar un sistema de exámenes colegiados que evaluarán los aprendizajes alcanzados por los alumnos al inicio, transcurso y término de sus estudios, mediante la selección de materias claves o integradoras que sirvan de parámetro para evaluar la calidad de los aprendizajes logrados.

- Integrar en una base de datos la información disponible de los indicadores del proceso de aprendizaje, de tal manera que se pudieran comparar entre ellos y consigo mismos a través del tiempo.
- Adoptar una política de transparencia respecto a la información contenida en la base de datos, de modo que tanto las autoridades educativas como los estudiantes, los académicos y el público en general, tuvieran acceso a la información de manera directa.
- Realizar ajustes en los programas y servicios educativos a fin de mejorar la calidad de los mismos, con base en los resultados proporcionados por la evaluación.
- Llevar a cabo estudios que dieran seguimiento al mejoramiento de la calidad de la educación, a partir de los indicadores.

Producto de tales intenciones, en marzo de 2004 la UABC publicó una convocatoria para la elaboración de exámenes colegiados que evaluaran el aprendizaje de los estudiantes. Dicha convocatoria mencionó dentro de sus bases “que el seguimiento educativo debe ser un proceso permanente, que la evaluación del aprendizaje, el fortalecimiento de la vida colegiada y las alternativas académicas para organizar el aprendizaje desde perspectivas innovadoras, abiertas y flexibles posibilitan el fortalecimiento del modelo educativo...” En dicha publicación, se convocó a directivos, cuerpos colegiados y académicos de la UABC a participar en el diseño de instrumentos de evaluación colegiada del aprendizaje que contribuyeran al

mejoramiento continuo de la calidad del aprendizaje (Universidad Autónoma de Baja California, 2004).

Por su parte, en el año 2003 el Colegio de Bachilleres de Baja California inició un proceso de elaboración de exámenes sumarios para certificar el aprendizaje que lograron sus estudiantes al final de los cursos que integran el plan de estudios. De hecho, el COBACH contaba ya con experiencia institucional para desarrollar esta clase de instrumentos, pues hasta finales de los años ochenta del siglo pasado desarrolló durante varios años exámenes departamentales para evaluar y certificar el aprendizaje de sus estudiantes.

En ese mismo año, los Colegios de Estudios Científicos y Tecnológicos de Baja California se dieron también a la tarea de construir exámenes para monitorear la calidad de los aprendizajes que lograron sus alumnos tras estudiar los cursos que integraron las principales materias y áreas de su plan de estudios.

Cabe señalar que, en virtud de que sus investigadores contaban con una experiencia reconocida para elaborar esta clase de instrumentos, las dos instituciones de bachillerato mencionadas solicitaron la colaboración del Instituto de Investigación y Desarrollo Educativo (IIDE) de la UABC para desarrollar sus respectivos exámenes.

Esta dependencia estaba en condiciones de atender tales solicitudes en virtud de que, entre fines del siglo pasado y principios del actual, el IIDE adaptó y desarrolló un modelo psicométrico para evaluar el aprendizaje que fue propuesto originalmente

por Anthony Nitko (1994) para desarrollar pruebas criterioles de gran escala alineadas con el currículum (Contreras, 2000).

Así, en septiembre de 2003 el IIDE inició un proyecto de evaluación educativa denominado *“Capacitación y asesoría a directivos y docentes del Colegio de Bachilleres de Baja California y del Colegio de Estudios Científicos y Tecnológicos de Baja California, para desarrollar y validar exámenes para evaluar el aprendizaje de los estudiantes”*. En dicho proyecto, se buscó capacitar en una primera fase totalmente presencial, a los directivos, académicos y docentes de ambas instituciones en el manejo de elementos básicos de psicometría que les permitieran desarrollar instrumentos de evaluación a gran escala del aprendizaje de sus alumnos. Se trabajó con tres tipos de comités para la elaboración de cada examen: el diseñador de cada instrumento, el elaborador de las especificaciones de ítems y el elaborador de los reactivos y modelos de examen. Estos comités estuvieron conformados por docentes de cada institución. La estructura del modelo estableció que cada comité desarrollaría una tarea específica en la cual generaría productos que servirían de base para que el siguiente comité continuara el trabajo de construcción del instrumento. El resultado de este proyecto fue la elaboración de siete exámenes sumarios por parte del COBACH, correspondientes a los cursos del primer semestre de su plan de estudios. Por su parte, el CECyTE desarrolló cinco pruebas para monitorear la calidad de los aprendizajes que logran sus alumnos al

finalizar los cursos correspondientes a otras tantas materias estratégicas del plan de estudios.

En el primer semestre de 2004 la operación del modelo continuó, pues el COBACH solicitó de nueva cuenta apoyo en la capacitación y asesoría necesarias para la elaboración de los siete exámenes sumarios correspondientes al segundo semestre del plan de estudios. En esa ocasión, se trabajó una sesión intensiva de capacitación presencial con cada tipo de comité y se proporcionó asesoría presencial a cargo de un académico del COBACH que participó en el proyecto anterior, mismo que realimentó los trabajos del comité respectivo, recopiló los productos y los envió al IIDE para ser revisados y realimentados una vez que contaron con el consenso del comité.

Finalmente, la colaboración con el COBACH continuó en el segundo semestre de 2004, cuando se implementó el modelo en las modalidades presencial intensiva y a distancia, en donde la capacitación que recibieron los diferentes comités fue llevada a cabo por personal de su institución que estuvo inmerso desde el proyecto inicial, y quien ya contaba con la experiencia teórico-práctica en la creación de exámenes. La revisión y realimentación finales de los productos generados por cada comité fue realizada a distancia por especialistas de la UABC.

Simultáneamente, en 2004 la operación del modelo se puso a prueba en un nuevo contexto, debido a que en la UABC se creó un programa institucional para la elaboración de exámenes para la evaluación colegiada del aprendizaje en la

institución. Con ello se vio la necesidad de adaptar el modelo a las necesidades propias de esta universidad y a la naturaleza y dimensiones de los grupos docentes en la entidad que requerirían capacitación y asesoría en forma simultánea.

Este proyecto se operó en una modalidad semipresencial y a distancia, similares a las que se trabajaron con el COBACH. Sin embargo, se incluyeron nuevas dimensiones a la operación del modelo: los manuales técnicos, formatos, ejemplos y otros materiales desarrollados se alojaron en una página Web para apoyar la capacitación y asesorías que fueron proporcionadas de manera presencial, semipresencial y a distancia, vía un foro asincrónico y un chat creados *ex profeso*.

1. 2. Planteamiento del problema

Tras varios años de desarrollo, el modelo psicométrico en cuestión ha producido una documentación considerable, se ha implementado con finalidades diversas, ha operado en diferentes condiciones, con distintas poblaciones y los procedimientos de capacitación y asesoría han sido adaptados para satisfacer un amplio rango de necesidades institucionales. No obstante, el modelo dista mucho de estar consolidado. Hay varias áreas de trabajo que aún no se han desarrollado y existen fases y procedimientos que contempla el modelo original que todavía no se trabajan. Por ejemplo, se cuenta con pocos datos empíricos derivados de la aplicación a gran escala de los instrumentos elaborados hasta el momento, que permitan realimentar la metodología mediante la que fueron desarrollados los ítems y

modelos; tampoco existe una documentación o registro de la experiencia educativa de la aplicación del modelo, ni se ha efectuado un proceso de evaluación formal de la operación del mismo, a pesar de que ha sufrido numerosos cambios al adaptarse a diversas condiciones.

Una notable limitación del modelo radica en el hecho de que, si bien contempla la capacitación, asesoría y realimentación de los trabajos de los comités de examen que coordinan la construcción de las pruebas, de quienes las diseñan, de los que elaboran las especificaciones de ítems y de quienes finalmente elaboran los reactivos y los calibran, una vez que los modelos son estructurados y se aplican a gran escala, no se han desarrollado procedimientos o materiales específicos para proporcionar un entrenamiento, asesoría y realimentación dirigidos a los responsables institucionales de los exámenes para analizar los datos derivados de su aplicación o para elaborar informes especializados sobre los resultados de los exámenes para los distintos usuarios de la información obtenida.

Así, cada institución define de manera más o menos intuitiva qué datos le interesa procesar y analizar y el tipo de informe de resultados que elaborará para documentar los resultados de la aplicación de los exámenes. Sin embargo, por comentarios aportados por las instituciones con las que se ha colaborado, se sabe que los análisis de datos así realizados tienden a ser los más simples y directos; usualmente solo aquéllos que permiten calificar los exámenes y asignar las notas correspondientes a los estudiantes y elaborar un informe de resultados que contiene

una estadística muy general sobre porcentajes de acreditación y reprobación por plantel y materia.

En consecuencia, el análisis de la información lograda mediante la aplicación de los instrumentos y la subsiguiente elaboración de los informes de resultados, no resultan todo lo técnicos y significativos que podrían y debieran ser, sobretodo si se considera la enorme inversión en recursos humanos, financieros y tiempo, efectuada por las instituciones que elaboran pruebas de gran escala.

Lo anterior plantea la necesidad de desarrollar un proceso explícito que incluya procedimientos y materiales para proporcionar capacitación, asesoría y realimentación de productos elaborados, que suministre los elementos técnicos básicos para que los diferentes usuarios del modelo puedan analizar, interpretar y generar informes significativos a partir de los resultados y la información obtenida tras la aplicación a gran escala de los instrumentos de evaluación diseñados.

En vista de lo anterior, se estima que el presente trabajo puede contribuir a lograr la formalización del proceso de análisis de resultados y elaboración de informes, así como aportar elementos que permitan establecer vínculos de comunicación entre usuarios a través de informes técnicos especializados que aporten avances en la investigación de la evaluación del aprendizaje que se realiza en nuestro Estado y en el país.

1. 3. Objetivos

El presente estudio tiene como propósito general:

1. Desarrollar procedimientos, instrumentos y materiales para capacitar y asesorar a los responsables institucionales de las pruebas criteriosales desarrolladas para evaluar el logro educativo, en el análisis de datos y la elaboración de informes técnicos de resultados derivados de la aplicación a gran escala de los instrumentos.

- **Objetivos específicos**

- 1.1. Diseñar los procedimientos necesarios para efectuar el análisis de datos derivados de la aplicación de pruebas criteriosales de gran escala y la elaboración de informes especializados de resultados según las necesidades de información de los usuarios de la misma.
- 1.2. Elaborar un manual técnico, formatos especializados y ejemplos de aplicación para apoyar la capacitación, asesoría y realimentación de productos del análisis de datos derivados de la aplicación de pruebas criteriosales de gran escala y de los procedimientos para la elaboración de informes de resultados.
- 1.3. Diseñar instrumentos para evaluar la operación de los procedimientos y materiales elaborados.

1. 4. Justificación

Existen varias razones que justifican realizar el presente estudio:

- El modelo para la construcción de pruebas criterioles de gran escala que se ha desarrollado en el IIDE durante más de diez años y que se ha utilizado para apoyar la elaboración de exámenes para la educación básica, media superior y superior, hasta el momento no cuenta con materiales y procedimientos para apoyar a los responsables institucionales de los exámenes, en el análisis de la información que se obtiene después de aplicar a gran escala los modelos de examen contruidos.
- De igual modo, el modelo no dispone aún de materiales y procedimientos para apoyar a los responsables institucionales de los exámenes, para que elaboren informes especializados del ítem análisis¹ efectuado ya con suficientes datos empíricos o para que construyan informes de resultados del examen por área del conocimiento y destinatario de la información, una vez que estos se obtienen tras aplicar a gran escala los modelos de examen elaborados.
- Hasta el momento, la transferencia de la tecnología contenida en el modelo para elaborar pruebas criterioles de gran escala, se ha llevado a cabo mediante una capacitación y asesoría proporcionadas mediante tres módulos cuyos focos han sido, respectivamente, el diseño de las pruebas, la elaboración de especificaciones de reactivos y la elaboración de ítems, así

¹ Para objeto de este documento, a partir de este momento se utilizará el término “análisis de reactivos” para referirse al ítem análisis.

como su posterior estructuración en modelos de examen y calibración psicométrica. Sin embargo, la transferencia de esta tecnología resulta visiblemente incompleta y objetable si se deja a los responsables institucionales de los exámenes sin una capacitación y asesoría formales y especializadas para que estén en condiciones de analizar los resultados de sus exámenes y elaborar los correspondientes informes para los destinatarios institucionales de la información. En consecuencia, se requiere un módulo adicional, centrado en el análisis de datos y la elaboración de informes de resultados de las pruebas, mismo que se desarrolla en el presente trabajo.

1. 5. Supuestos

Se considera que la creación y evaluación de un módulo especializado para llevar a cabo la capacitación, asesoría y realimentación necesarias para que los responsables institucionales de los exámenes elaborados estén en condiciones de analizar los resultados que se obtienen después de aplicarlos a la población objeto, así como para que puedan construir los informes especializados que den cuenta, de manera diferencial, de la ejecución que tuvieron los examinandos en las pruebas, no solo constituye un complemento deseable y necesario del modelo para desarrollar pruebas criterioles de gran escala, sino que de hecho permitirá que los responsables del modelo puedan asumir plenamente el compromiso de transferir a las instituciones

que lo utilizan, la metodología para elaborar esta clase de instrumentos y sacar el debido provecho tras su aplicación.

También se considera que, aunque no ha sido debidamente documentada, la experiencia acumulada durante el desarrollo y operación del modelo psicométrico para construir pruebas criterioles de gran escala, así como la información aportada por las instituciones educativas que lo han utilizado para construir sus pruebas y aplicarlas a sus estudiantes, son suficientes para orientar el diseño de los procedimientos y materiales que se incluyen en el módulo que se elaboró.

Capítulo II.- ANTECEDENTES TEÓRICO-CONCEPTUALES Y CONTEXTO PSICOMÉTRICO DE LA EVALUACIÓN DEL APRENDIZAJE

2. 1. Aspectos conceptuales generales sobre la evaluación psicométrica.

A lo largo de la historia, se han contemplado diversas visiones de la medición psicológica. Sus funciones han sido modificadas y adaptadas según las necesidades de la sociedad y los cambios surgidos dentro de ella. Se cuenta con algunos datos acerca de esta clase de medición desde Aristóteles y Platón, quienes escribieron sobre las diferencias individuales hace aproximadamente 2500 años; pero se tienen antecedentes inclusive desde el año 2200 A. C., cuando los Chinos aplicaban rigurosos sistemas de examinación para seleccionar, promover y retener oficiales del servicio público, en los cuales se utilizaban algunos principios de la medición (Ward, Stoker y Murria, 1996).

No fue sino hasta el siglo XIX que se iniciaron los estudios de las diferencias entre individuos en cuanto a sus capacidades mentales y personalidad (Aiken, 1996). Expresamente se considera que el primer laboratorio de psicología experimental, fundado por Wilhem Wundt en 1879, fue donde se realizaron los primeros estudios formales en los cuales el ser humano fue objeto de medición. Al finalizar el siglo XIX, Francis Galton, se interesó especialmente por la herencia del talento y se le considera padre del análisis estadístico (Ward, Stoker y Murria, 1996); incursionó en el desarrollo de escalas de puntuación y cuestionarios, siendo una de sus

contribuciones metodológicas más importantes la técnica de correlaciones. Posteriormente, Cattell utilizó los aportes de Galton y correlacionó calificaciones de pruebas mentales de tiempo de reacción y discriminación sensorial con calificaciones escolares. Pero fue Alfred Binet quien en 1905 elaboró la primera prueba mental que probó ser un indicador efectivo del logro académico.

En lo que respecta al siglo XX, hay varios eventos que han marcado la pauta en el desarrollo de la psicometría. En primer lugar, la primera guerra mundial creó la necesidad de desarrollar una forma de evaluar a grandes masas para distinguir las capacidades mentales de los soldados de las fuerzas armadas estadounidenses. Así fue como surgió el test Army Alfa, que se aplicaba a quienes sabían leer y escribir; y el Army Beta, para los analfabetas. Posteriormente, a partir de la década de los años 30 se desarrolló sustancialmente la teoría psicométrica y se crearon estándares que incluían los lineamientos para la construcción, evaluación, documentación, usos y aplicaciones de las pruebas, así como la generación de informes de resultados a distintos destinatarios (APA, 1999).

2. 2. Tipos de pruebas que evalúan el aprendizaje

En el ámbito de la evaluación del aprendizaje a gran escala existen dos tipos de pruebas: las que interpretan la ejecución de los estudiantes en relación a una norma, y las que lo hacen en relación a un criterio.

Las primeras se denominan pruebas referidas a una norma debido a que se hace una comparación de los puntajes obtenidos por los examinandos en relación a otros de su mismo grupo; a esta población se le denomina grupo normativo. En las pruebas referidas a un criterio se interpretan los puntajes de un estudiante en relación al nivel de dominio que éste tiene respecto a un criterio definido; por ejemplo, los segmentos del conocimiento que se espera haya aprendido como resultado de la operación curricular.

Las pruebas criterioles tienen varios nombres: referidas a un dominio, de dominio, de competencia, de habilidades básicas, o exámenes de certificación.

La distinción de ambos tipos de pruebas por su referencia, significa que no es posible interpretar los puntajes crudos obtenidos por los estudiantes en la prueba sin referirlos a una norma o criterio. También es importante señalar que estos tipos de evaluación se requieren para interpretar los puntajes individuales de manera válida (Nitko, 1994).

Glasser (citado por Hambleton, 1988), publicó en 1963 un artículo que menciona algunos métodos de evaluación que proporcionan la información necesaria para la toma de decisiones en relación a objetivos y competencias específicas. Por su parte, Popham (1978, citado por Hambleton en 1988), fue quien introdujo una definición de evaluación criterial que hasta el momento es la más utilizada en el ámbito educativo. Menciona que el objetivo de las pruebas referidas a un criterio es permitir la interpretación de la ejecución de los examinandos en relación a

competencias que previamente fueron establecidas. Esto es de suma importancia, ya que así es más fácil elaborar ítems que correspondan en forma directa a dichas competencias, así como llevar a cabo una interpretación adecuada de los resultados de la ejecución de los examinandos.

Por ello, debe establecerse un estándar para cada criterio o área que mida la prueba, ya que se evaluará la ejecución de los examinandos en relación a dichos criterios y posteriormente se llevará a cabo una interpretación de los resultados de la prueba. En este sentido, la información que proveen las pruebas criterioales es de gran valor para los destinatarios.

Cabe mencionar que los resultados de estas pruebas tienen tres principales usos (Hambleton, 1988):

- 1) Describir la ejecución de los examinandos
- 2) Determinar el nivel de dominio de los examinandos (diferenciar entre aquellos alumnos que tienen el nivel de maestría y los que no lo poseen), y
- 3) Describir la ejecución de grupos específicos de examinandos, dentro de estudios de evaluación de programas.

Asimismo, existen tres aspectos importantes a considerar en una prueba criterial:

- 1) La cantidad de criterios evaluados varía de prueba a prueba
- 2) El número de ítems que mide cada criterio y el valor mínimo estándar va a variar de un criterio a otro; y

- 3) Para establecer si el examinando domina o no el criterio, se debe comparar la proporción de aciertos correctos con el estándar de ejecución mínimo.

Existen pruebas referidas a un criterio o dominio, y las referidas a un objetivo. La diferencia entre unas y otras es que las primeras organizan los ítems en grupos que evalúan contenidos bien definidos, y que a la vez son representativos de contenidos específicos que miden un objetivo; mientras que en las pruebas referidas a un objetivo no se especifica un dominio de contenido, y los ítems no son considerados como representativos de ningún contenido. Por ende la interpretación de la ejecución de los examinados en las pruebas referidas a un objetivo debe circunscribirse a los ítems específicos de la prueba (Hambleton, 1988).

Las pruebas normativas son necesarias para proveer datos normativos válidos y confiables que sirvan para comparar a los examinados entre sí. Las pruebas criterioles facilitan la interpretación de la ejecución de los examinados en relación a competencias bien definidas.

De esta manera se puede decir que el uso de pruebas referidas a una norma tiene mejores resultados cuando se desea medir un área amplia del conocimiento; mientras que las pruebas referidas a un criterio se utilizan cuando lo que se busca es evaluar la competencia de manera específica dentro de un dominio de conocimiento (Klein, 1990).

2. 2. 1. Similitudes y diferencias entre pruebas normativas y pruebas criterios

Aunque ambos tipos de pruebas difieren en la forma en que son construidas y en la manera en que se interpretan sus resultados, hay una serie de similitudes y diferencias entre ambos tipos de pruebas que a continuación se revisarán. Existen tres principales similitudes:

- a) Ambas pruebas utilizan el mismo formato de elaboración de ítems
- b) Las instrucciones en ambos casos son similares
- c) Los dos tipos de pruebas se pueden estandarizar.

En la siguiente tabla, se encuentran de manera detallada una descripción de cada uno de los aspectos que marcan las principales diferencias entre ambos tipos de pruebas (Hambleton, 1988; Gómez, 1990).

Tabla 1. Diferencias entre pruebas referidas a una norma y referidas a un criterio

CARACTERÍSTICA	PRUEBAS NORMATIVAS	PRUEBAS CRITERIALES
Propósito de la prueba	Estas pruebas son construidas específicamente para facilitar la comparación entre los examinandos dentro del área contextual que mide la prueba. Es muy común que se utilicen normas percentilares de edad, o puntajes normativos estandarizados. Se pueden hacer interpretaciones basadas en la ejecución de los examinandos, pero debido a que la cantidad de ítems que miden las diferentes competencias de las pruebas es poca, el valor de dichas interpretaciones se encuentra limitado.	Las pruebas referidas a un criterio son construidas para evaluar la ejecución de los examinandos en relación a una serie de competencias. Los puntajes pueden ser utilizados para describir la ejecución del examinando, para determinar si domina o no el contenido, y para evaluar la efectividad de los programas educativos. En el caso de estas pruebas el puntaje puede utilizarse para comparar a los examinandos, tomando en cuenta que la confiabilidad será menor si la distribución de puntajes resulta ser homogénea.

CARACTERÍSTICA	PRUEBAS NORMATIVAS	PRUEBAS CRITERIALES
Especificaciones de los contenidos a evaluar	El grado de detalle con el que son elaboradas las especificaciones de la prueba no requiere ser muy fino.	En estas pruebas se debe elaborar una serie de especificaciones mucho muy detalladas, de tal forma que se asegure que los resultados de los examinandos sean interpretados debidamente.
Desarrollo de la prueba	El análisis estadístico de los ítems es muy importante para hacer una adecuada selección de los ítems de la prueba. Los que presentan un grado de dificultad moderado (entre .30 y .70), así como un alto poder de discriminación (correlación r_{Bis} mayor a .30) son los que tienen más probabilidad de ser seleccionados para la prueba, ya que contribuyen sustancialmente a la varianza del puntaje. La confiabilidad y validez son más altas cuando la varianza aumenta.	En este tipo de pruebas se hace una eliminación de los ítems solamente cuando se determina que violan los principios establecidos en las especificaciones, o cuando el análisis estadístico de los ítems revela fallas que no se pueden corregir.
Generalizabilidad del puntaje de la prueba	Existe muy poco interés en hacer generalizaciones basadas en los resultados de la prueba.	Normalmente se tiene mucho interés en la generalización de los puntajes de la prueba, ya que cuando hay especificaciones claras, y los ítems son representativos del dominio de contenido a evaluar, se puede generalizar la ejecución del examinando a dominios del contenido más amplios.
Uso e interpretación de los resultados	Los resultados del examinando son una muestra del nivel general de funcionamiento, más que del dominio de contenidos, habilidades o información específica alcanzada en un área, y son muy útiles cuando todos los examinandos dominan el contenido en determinado nivel, pero se desea saber quienes lo dominan a niveles superiores.	En estas pruebas los resultados son una muestra del nivel de competencia del examinando en un repertorio de habilidades bien definido, las cuales se fijaron como objetivos a dominar y son útiles cuando se busca medir el nivel de aprendizaje adquirido a lo largo y al final del proceso educativo.

2. 2. 2. Determinación del tipo de prueba que se utilizará

Woolfolk y McCune (1983) proponen tomar en cuenta las siguientes características al decidir el tipo de prueba que se va a utilizar:

- 1) Las pruebas normativas son mejores para medir capacidades generales de los alumnos en áreas específicas como las ciencias, historia, lenguas, etc.; sirven para estimar la gama de capacidades de un grupo grande y para hacer una selección de los mejores candidatos cuando los lugares son escasos.
- 2) Las pruebas criterioles son mejores para medir el dominio de destrezas básicas, para determinar si el alumno cumple los requisitos previos para pasar de unidad o grado, para estimar objetivos afectivos y psicomotores, y para agrupar a los alumnos para la instrucción, basándose en lo que saben y también en lo que necesitan aprender.

Nitko (1994) señala que es muy importante considerar que el diseño de un instrumento de evaluación debe ser consistente con las intenciones de los educadores y las metas de los sistemas educativos en dónde van a ser utilizados dichos instrumentos. De no tomar en cuenta esto, existe la posibilidad de que se presenten una serie de consecuencias importantes. La primera de ellas implica que tanto docentes como estudiantes estarán mal informados respecto a las dificultades y avances que tengan los alumnos. La segunda señala que se puede ver reducida la motivación de los estudiantes para aprender. Otra consecuencia es que puede haber una inadecuada toma de decisiones críticas, tal como sería el caso de decidir otorgar

o no un título a un alumno. La última consecuencia que menciona el autor es que puede hacerse una evaluación incorrecta de la efectividad instruccional.

2. 2. 3. Uso de las pruebas criterioles y de sus resultados en la elaboración de informes de resultados

Actualmente se utilizan los resultados de las pruebas criterioles en áreas muy diversas. Por ejemplo, los maestros, utilizan los resultados de la ejecución de los alumnos para ubicarlos en programas educativos adecuados a sus necesidades, monitorear sus avances, realimentar su aprendizaje, identificar sus deficiencias y también para planear su instrucción. En el área de educación especial este tipo de informe ayuda mucho en el diagnóstico de problemas de aprendizaje, así como en el monitoreo del progreso de los alumnos. También se utiliza este tipo de informes para evaluar programas educativos normales y avanzados. Pero quizás uno de los usos más importantes que se le da a los resultados de las pruebas se ubica en el área de certificación profesional (Hambleton, 1988).

En el ámbito internacional, en 1994 el Congreso de los Estados Unidos aprobó el Acta de las Escuelas de América documento en el cual se estipula en el Título I, como una obligación, y a la vez como una oportunidad para los estados, organismos educativos y escuelas, desagregar y analizar los datos resultantes de las evaluaciones de sus estudiantes (Jaeger y Tucker, 1998). Estas evaluaciones permitirán la desagregación de la información en variables específicas, tales como

género, escuela, desempeño en el área de matemáticas, estatus migratorio, discapacidades, situación económica, etc. Sin embargo, esto implica una necesidad de definir adecuadamente dichas variables. Entre otras cosas, es necesario especificar definiciones de las categorías de los estudiantes, de los aspectos legales en cuanto a la confidencialidad de los resultados, de la generalización adecuada de los resultados de la evaluación, de la interpretación válida de los resultados de la evaluación dentro y entre las categorías de los estudiantes, y de la difusión de los resultados de tal forma que sean útiles a distintos destinatarios.

Lo anterior brinda una gran oportunidad para llevar a cabo un análisis nuevo y creativo de la información, cuyo uso podría extenderse al análisis de las políticas educativas actuales, evaluación de los programas existentes, planeación y desarrollo de intervenciones instruccionales innovadoras, así como la elaboración de informes respecto al estado actual y al progreso de los alumnos. Quienes formen parte activa del sistema educativo y tengan injerencia en el análisis y reporte del desempeño colectivo de los alumnos, se verán beneficiados al desagregar la información que arrojan los resultados de las evaluaciones. Dentro de esa población se encuentran maestros, directivos, inspectores de zona, administradores, instituciones de educación y organismos dedicados a la evaluación educativa, e inclusive agencias gubernamentales que toman decisiones en torno a políticas educativas a nivel local, estatal o nacional (Jaeger y Tucker, 1998).

2. 3. Normatividad internacional y nacional para la elaboración, evaluación y uso de pruebas

En 1951 se creó la primera edición del libro *Educational Measurement*, cuyas ediciones posteriores hasta la fecha lo han convertido en un verdadero clásico en el campo de la evaluación educativa. A partir de esta publicación, en 1954 la American Psychological Association (APA), la American Educational Research Association (AERA) y el National Council on Measurement in Education (NCME), se constituyeron en un comité conjunto para normar las prácticas relativas a la elaboración y uso de instrumentos de evaluación psicológica. Así, publicaron los primeros estándares internacionales para tests educativos y psicológicos que mencionan los criterios técnicos que norman la construcción, evaluación y documentación de pruebas (APA, 1999). Entre dichos criterios, los más importantes se refieren a la validez, confiabilidad y errores de medición, desarrollo y revisión del instrumento de evaluación, administración, calificación, y generación de informes de resultados, uso justo y adecuado de la evaluación, así como las responsabilidades de los usuarios del instrumento.

En México, tras su creación en 1994, el CENEVAL consideró necesario someter sus procesos y productos psicométricos a evaluaciones externas rigurosas que se apegaran a los estándares de calidad más elevados. Para ello, constituyó un Consejo Asesor Externo que desarrolló los primeros estándares de calidad para la elaboración, aplicación y uso de instrumentos para la evaluación educativa,

publicados en y para nuestro país (Consejo Asesor Externo del Centro Nacional de Evaluación para la Educación Superior, 2000). Para la generación de este documento se utilizó como punto de partida el cúmulo de experiencias en la práctica del campo de la evaluación educativa resumidas en los estándares publicados por la APA-AERA-NCME en 1981, 1985, 1994 y 1997.

Una vez que se diseña, elabora y administra un instrumento de evaluación, basado en los estándares establecidos por el comité conjunto APA-AERA-NCME, resulta necesario llevar a cabo un análisis y una devolución de resultados que impliquen una adecuada interpretación de los mismos, para así generar informes que cumplan los propósitos de la prueba, inicialmente marcados por sus diseñadores. Para ello debe considerarse a qué tipo de población se dirige la información, el tipo de lenguaje que resultará claro y apropiado, y que reflejará el significado y utilidad que tienen los resultados dentro del contexto de la evaluación. En este sentido, todo instrumento de evaluación diseñado debe estar debidamente acompañado por una serie de documentos que ayuden al usuario a contextualizarse y dar un uso adecuado a la información. Dichos documentos podrán ser manuales técnicos, guía para el usuario, instrucciones para el administrador y evaluador, guía de estudios para el alumno, etc., los cuales deberán incluir dentro de su información algunos datos como la naturaleza de la prueba, su propósito y el proceso de elaboración; información técnica sobre su administración, indicaciones para la calificación e interpretación de resultados; información respecto a las evidencias que sustentan su

validez y confiabilidad; normas de calificación, y otra información que provea a los usuarios con todo aquello que necesiten para poder juzgar la calidad del instrumento, de sus resultados y del uso que se dará a los mismos (APA, 1999).

En esta misma línea, los Estándares establecidos para nuestro país por el Consejo Asesor Externo del CENEVAL mencionan en forma más específica algunos aspectos relativos a la comunicación de resultados; tales como la importancia de reportarlos de manera que todas las partes involucradas los reciban en un lapso razonable, que puedan interpretarlos correctamente, y que incluyan información del puntaje obtenido, datos normativos, criterios de referencia o puntos de corte. También mencionan que se deberá incluir una descripción de los propósitos y características de la prueba, lo que mide y no mide, las conclusiones y decisiones que pueden basarse en los resultados, y todo aquello que apoye a la correcta interpretación de los mismos (Consejo Asesor Externo del Centro Nacional de Evaluación para la Educación Superior, 2000).

En el caso de los Organismos Dedicados a la Elaboración de Pruebas (ODEP), los Estándares mencionan que además de la información individual de los resultados de la evaluación, deberán hacer la difusión de material de apoyo dirigido a las partes involucradas en la evaluación, con la finalidad de que se haga un uso correcto de los resultados, y de que éstos sean insumos para la toma de decisiones pertinentes. Lo anterior implica que los ODEP tienen la responsabilidad de promover el aprovechamiento de los resultados de las evaluaciones, tanto en las instituciones

donde se realice la evaluación, como en las de niveles o ciclos anteriores y subsiguientes; así como en los sistemas educativos estatales y nacional. Así, los Estándares señalan que los ODEP deberán ofrecer oportunidades de formación para interpretar correctamente los resultados de las pruebas en los organismos usuarios, ya que estas instancias son elementos valiosos de retroalimentación para los procesos de revisión curricular y para el monitoreo de la calidad educativa, y porque pueden contribuir al diseño y establecimiento de políticas y mecanismos de mejoramiento más adecuados (Consejo Asesor Externo del Centro Nacional de Evaluación para la Educación Superior, 2000).

En cuanto a estándares nacionales relacionados con la capacitación, en 1995 se creó en México el Proyecto de Modernización de la Educación Técnica y la Capacitación (PMETYC), cuyo objetivo fundamental es: "... impulsar un proceso de cambio capaz de convertir la formación de los recursos humanos en el eje central del progreso personal y profesional de los trabajadores, así como del aumento de la productividad y la competitividad de las empresas mexicanas" (Fraustro, 2000 p. 52). En dicho proyecto se propone que sea el sector productivo quien norme los niveles de competencia requeridos en el campo laboral, a través del Consejo de Normalización y de Certificación de Competencia Laboral (CONOCER), integrado por representantes del sector empresarial, social y secretarías de Estado. Los principales propósitos del CONOCER son planear, operar, fomentar y actualizar los sistemas de normalizado y de certificación de competencia laboral que forman parte del

PMETYC. En este sentido, en México, a través de esta política nacional, se deben seguir ciertos estándares o normas técnicas para el diseño e impartición de cursos de capacitación, tales como la elaboración de un manual del participante y otro para el instructor, así como preparar, conducir y evaluar el proceso de capacitación. Dichos estándares dictaminan la forma en que se debe estructurar un programa de capacitación, y por ende los manuales de procedimientos y demás materiales didácticos a utilizar en el proceso (Consejo de Normalización y Certificación de Competencia Laboral, 2002).

2. 4. Modelos psicométricos para evaluar el aprendizaje

Durante los últimos 25 años se han desarrollado diversos modelos psicométricos para construir y validar instrumentos destinados a evaluar el logro educativo, los cuales pueden ser clasificados en dos grandes categorías: los que se encuentran orientados a elaborar pruebas referidas a una norma, y los que buscan diseñar pruebas referidas a un criterio.

Como ya se señaló, con las pruebas normativas se pretende interpretar la ejecución del examinando en relación con la de otros que también respondieron a la prueba. Así, se ubica su ejecución por arriba, abajo o dentro de lo normal o promedio, a partir de la medida que se toma de una muestra representativa o de toda la población de estudiantes. En consecuencia, este tipo de pruebas proporciona una medida de la habilidad relativa del estudiante. Por otro lado, las pruebas criterioles

interpretan la ejecución del examinando en relación a su estatus respecto a un criterio o dominio evaluativo bien definido y comunican lo que puede y no puede hacer el examinando en un campo de conocimiento, en relación a la medida que se toma de todo el dominio de contenido o de una muestra representativa de tareas obtenidas de ese dominio, usualmente con el propósito de hacer una adecuada evaluación de programas educativos, certificar competencias académicas, etc. (Berk, 1984; Nitko, 1984; Popham, 2000).

Los principales modelos para desarrollar pruebas criterioles fueron desarrollados a principios de los años ochenta del siglo pasado y quedaron plasmados en una guía para la construcción de pruebas referidas a un criterio, cuyo editor (Berk, 1984) compiló a partir de los trabajos de investigadores como James Popham, Anthony Nitko, Ronald Hambleton y otros, quienes posteriormente continuaron desarrollando sus propios modelos psicométricos para apoyar la evaluación del aprendizaje.

Entre estos modelos destaca el de Nitko (1994), quien propuso un modelo para construir exámenes nacionales referidos a un criterio orientado por el currículum, y en el cual establece que el desarrollo de una prueba criterial debe considerar como punto de partida al currículum, y que las decisiones respecto a qué evaluar y cómo hacerlo están determinadas por los resultados de aprendizaje que establece dicho currículum. El modelo plantea nueve etapas para construir el instrumento que incluyen:

- 1.- Definir el dominio de resultados que pretende el currículum. En esta primera fase se debe centrar la atención sobre los resultados que se pretenden obtener a través del currículum, de tal forma que el examen pueda quedar alineado con éste. Para ello deben revisarse y sintetizarse una variedad de fuentes de información, tales como: quién elaboró el currículum, en qué teoría cognitiva y curricular se sustenta, el contenido del programa educativo, las metas del currículum, las actividades que realizan los mejores maestros en el aula, y qué materiales instruccionales utilizan, entre otras.
- 2.- Analizar el currículum. En esta fase se hace un esquema o “mapeo” del currículum, de tal forma que se organizan sus contenidos y los resultados que se propone obtener a través de su instrucción, a fin de detectar aquellos que son particularmente relevantes.
- 3.- Desarrollar un plan de evaluación. Debe incluir la descripción de lo que se espera evaluar; una selección o muestreo del material a evaluar; la creación de tareas prototipo que tengan utilidad en diversos niveles educativos, cuidando que sean una representación fiel de la riqueza y variedad del currículum; publicación de las tareas; y creación de comités examinadores para cada una de las fases de desarrollo.

- 4.- Desarrollar la especificación de ítems. Esta tarea es importante debido a que en ella se describe la naturaleza de las tareas a evaluar, con suficiente detalle como para dejar claro a los elaboradores de ítems, cuál va a ser el dominio del conocimiento que deberá incluirse en la construcción de cada ítem.
- 5.- Producir y validar ítems. Esta fase tiene tres procedimientos que aseguran que el ítem es representativo del currículum. 1) La especificación de las tareas derivadas del currículum se utilizan como guía para la creación de ítems. 2) Cada ítem se revisa para verificar su apego a la especificación. 3) Debe hacerse un análisis empírico de los puntajes obtenidos por los estudiantes, para definir si el ítem funciona como debe, y si los examinandos lo interpretaron adecuadamente al responder.
- 6.- Ensamblar los exámenes. Se refiere a la elaboración y reproducción del examen, el cual debe representar al currículum de tal forma que evalúe con precisión el dominio del conocimiento que se supone que el alumno debe haber aprendido.
- 7.- Establecer estándares de ejecución. Debe asegurarse que éstos sean justos y representativos de la ejecución en forma comparativa de un año a otro.

8.- Realizar un análisis primario de los ítems elaborados. Esta fase es de especial interés para este estudio, ya que Nitko (1994) menciona que después de realizar una evaluación, los resultados deben analizarse y reportarse. También debe seguirse un mismo procedimiento para marcar o calificar las respuestas de los evaluados, para igualar los resultados del periodo actual con los de años anteriores, y para reportar los resultados de manera que resulte adecuada y significativa. Muchos programas de evaluación que existen a nivel nacional no llevan a cabo un análisis empírico completo de los datos cuando se reportan los resultados de la evaluación a los examinandos y a las autoridades educativas; sin embargo, aun cuando lo que se busca es publicar los resultados, es muy importante realizar un análisis de la calidad de la evaluación, inclusive cuando se crea una nueva versión de la misma el siguiente año.

En esencia, como parte de esta octava fase, debe implementarse y mantenerse un programa de control de calidad. Si se pretende crear evaluaciones criterioles de gran escala alineadas al currículum, deben especificarse y adoptarse estándares de calidad que se conviertan en la política oficial de los cuerpos o comités de examen que desarrollan el instrumento. Estos

estándares se encargarán de describir las propiedades técnicas que toda evaluación debe tener, y que a su vez la institución educativa requiere antes de seleccionar o certificar a sus estudiantes.

Existen tres áreas de control: contenido del ítem, calidad técnica de cada ítem, y calidad de las puntuaciones del test. La lista de estándares debe enunciar las cualidades que cada institución considera deseables para cada evaluación antes de que ésta sea de uso oficial y se tomen decisiones educativas a partir de ella. Cada estándar de calidad deberá contribuir en forma positiva a que la evaluación sea altamente válida y relevante para el propósito de certificación de competencias y selección de estudiantes para el nivel escolar superior inmediato.

- 9.- Realizar un análisis secundario para elaborar reportes o informes dirigidos a distintos tipos de usuarios. En esta última fase del modelo, se debe identificar a los destinatarios para que todos los informes que se generen de los resultados de la evaluación se diseñen de acuerdo a sus necesidades. En este grupo se puede incluir a las autoridades educativas, encargados de desarrollo curricular, inspectores escolares, directivos de escuela, etc.

En el caso de nuestro país, a fines del siglo pasado se adaptó el modelo de Nitko para desarrollar una prueba de español para la educación primaria en México,

reduciendo el número de etapas del modelo original, de nueve a seis para hacerlo viable y funcional en nuestro contexto. Dichas etapas incluyeron la definición del dominio de resultados que pretende el currículum, el análisis curricular para detectar el contenido importante a evaluar, desarrollar un plan de evaluación, producir y validar ítems, hacer un análisis primario para revisar la calidad técnica de los ítems y modelos, calificar la prueba y analizar sus resultados, y hacer un análisis secundario para generar informes de resultados (Contreras, 2000). Cabe señalar que el presente estudio forma parte de los esfuerzos que se están haciendo por continuar el desarrollo psicométrico del modelo adaptado por Contreras.

2. 5. Establecimiento de estándares de ejecución y puntos de corte

Los puntos de corte en una prueba deben establecerse con base en los objetivos que persigue el instrumento, particularmente en cuanto al tipo de referencia que utilizamos para interpretar los resultados. Así, las pruebas que utilizan puntos de corte establecidos de manera normativa usualmente persiguen el objetivo de seleccionar a los estudiantes más aptos y asignar recursos a gran escala (Contreras, 2000, pg. 15). Al establecer el punto de corte de manera normativa se utilizan porcentajes para interpretar los resultados de un estudiante en relación con las calificaciones de otros estudiantes, lo que constituye una interpretación relativa. Sin embargo, cuando se busca establecer puntos de corte para pruebas con propósito de certificación o de descripción de la ejecución en referencia al dominio de un criterio

como el currículum, se busca diferenciar a los examinandos en dos categorías: los que tienen dominio del conocimiento evaluado y los que no lo tienen. Por ello es necesario establecer estándares de ejecución, los cuales se obtienen únicamente a través de la implementación de puntos de corte de tipo absoluto (Payne, 1992, citado por Márquez, 2005).

Así, en las pruebas criterioles un punto de corte es una medida establecida en una escala de puntuación, utilizada para evaluar el desempeño de un examinando en relación a un dominio o competencia, y expresa el grado en que el examinando es o no competente para aplicar conocimientos, capacidades y habilidades adquiridas. En general, el punto de corte divide a los examinandos en dos grandes categorías: los que obtuvieron puntuaciones altas en la escala de calificación, llamados competentes; y los que obtuvieron una baja puntuación, denominados no competentes (Hambleton y Eignor, 1980; Jaeger, 1989; Popham, 1990).

Desde los años cincuenta han existido diversos métodos para establecer puntos de corte (Angoff, 1971; Ebel, 1972; Nedelsky, 1954; citados por Márquez, 2005). Dichos métodos generalmente se basaban en juicios emitidos por grupos de expertos, lo cual no resultaba totalmente confiable, ya que podía acreditar todo estudiante que demostrara tener conocimientos mínimos indispensables.

En el campo de la evaluación del aprendizaje existen dos modelos generales para establecer el punto de corte: los modelos estáticos y los continuos.

Los primeros consideran que el aprendizaje es dicotómico: se posee o no la competencia. Este modelo es funcional cuando el área o materia evaluada permite construir un examen con un elevado grado de homogeneidad en los reactivos, tal como el área de las matemáticas, ya que el punto de corte se asume como el 100% de las respuestas correctas.

Los modelos continuos se basan en las teorías de adquisición del conocimiento, donde se van presentando grados o niveles de dominio conforme se da el aprendizaje (Márquez, 2005).

Actualmente, algunos autores han buscado minimizar el factor de arbitrariedad del juicio de valor a través de la creación de más de 38 métodos, entre los cuales se pueden mencionar el de Berk, Hambleton, Livingston y Zieky, Jaeger y otros (Márquez, 2005). Berk, en 1986 propuso tres categorías para agrupar los procedimientos para determinar los puntos de corte: de jueceo, basado en opiniones que discuten grupos de expertos sin conocer el desempeño; de Jueceo-empírico, en donde los expertos sí tienen información acerca del desempeño; y el Empírico-Jueceo, donde se hace un análisis estadístico del desempeño de los examinandos y se define el criterio de dominio o no dominio. A continuación se encuentra una descripción de los modelos continuos que han causado mayor impacto a lo largo del tiempo, y que han sido utilizados más ampliamente para establecer puntos de corte. En cuanto al resto de los métodos, se puede decir que muchos de ellos son

modificaciones que surgieron a partir de los métodos clásicos que se describen en este estudio.

Método de Nedelsky (1954). Es un método de jueces que se utiliza con pruebas de opción múltiple. Se basa en el supuesto de que el examinando que posee una competencia mínima elimina en primer término las opciones de respuesta que a su juicio son incorrectas, y procede posteriormente a escoger la que considera correcta entre las que restan. Los jueces establecen qué opciones de cada ítem debería considerar erróneas el estudiante que tuviese la competencia mínima requerida para aprobar; posteriormente asignan el recíproco del número de opciones no seleccionadas por el examinando, de tal forma que si por ejemplo, en un ítem de cinco alternativas los jueces consideran que un alumno mínimamente competente debería saber que son falsas tres de las cinco, el valor que se busca sería 0.5 (recíproco de dos). Este proceso genera tantos recíprocos como ítems. Posteriormente se obtiene la suma de todos los recíprocos, y el valor resultante se interpreta como la puntuación que obtendría el alumno mínimamente competente para el juez en cuestión. El punto de corte se establece como la media de los valores resultantes producidos por los jueces (Abad y cols., s/f). Entre las fallas de este método se encuentra el hecho de que no se puede saber con certeza si el examinando utiliza el método de adivinación o no, además de que los valores asignados por los jueces regularmente quedan alrededor de 0.5, lo cual produce un punto de corte más bajo que otros métodos.

Método de Angoff (1971). También es un método de jueceo donde el grupo de especialistas revisa los resultados y estima la probabilidad de que un examinando pueda contestar correctamente la pregunta. Se suman las estimaciones de los especialistas y se obtiene el promedio que se utilizará para establecer el punto de corte. Es un método recomendado cuando se persiguen fines de certificación debido a que permite tomar decisiones en una línea continua de probabilidades, lo cual produce puntos de corte más directos y verosímiles; sin embargo, generalmente se obtienen variaciones grandes entre las probabilidades de los reactivos de cada juez, lo cual presenta una desventaja.

Método de Ebel (1972). Método de jueces donde se hace un análisis de la dificultad y relevancia de cada reactivo. Implica un procedimiento de dos etapas. Cada juez clasifica primero las preguntas en grupos y luego emite un solo juicio numérico para cada grupo de preguntas. La clasificación de preguntas en grupos se basa en dos tipos de juicios sobre cada pregunta: un juicio de su dificultad y un juicio de su relevancia (o importancia). Ebel sugirió tres niveles de dificultad, denominados “fácil,” “mediano” y “difícil”; y cuatro categorías de importancia, denominadas “esencial,” “importante,” “aceptable” y “objetable.” La primera tarea del juez es clasificar todas las preguntas del examen. La segunda tarea del juez es emitir juicios sobre el rendimiento de un examinando que posee una competencia mínima. El juez debe emitir un juicio de este tipo por cada uno de los 12 bloques de la tabla de clasificación (excepto los que estén vacíos). Es decir, debe emitir un juicio para las

preguntas clasificadas “esenciales, fáciles”; otro para las preguntas clasificadas “esenciales, medianas”; y así sucesivamente hasta llegar a “objetable, difícil.” El juicio consiste en una respuesta a la pregunta: “Si un examinando con competencia mínima tuviera que contestar un gran número de preguntas como éstas, ¿Qué porcentaje contestaría correctamente?”. Tal como se mencionó, este método permite a los jueces tomar en cuenta la dificultad y la importancia de cada pregunta del examen. Esta característica es especialmente valiosa cuando las preguntas del examen difieren ampliamente en importancia. Sus desventajas son su lentitud y su inadecuación para exámenes cortos (Livingston y Zieky, 2004).

Método de Jaeger (1978). Este método está clasificado en la categoría de Jueceo-Empírico y se considera interactivo porque los jueces que participan tienen diversos marcos de referencia, y utiliza información normativa del desempeño de los examinandos. En lugar de buscar competencias mínimas, se busca si el examinando puede o no contestar correctamente el reactivo, y con ello determinar si se le otorga o no el documento que acredite su conocimiento. El punto de corte es el cálculo del promedio de las respuestas de los jueces. El proceso interactivo de este método permite reducir la variabilidad en la distribución de los estándares recomendados, y con ello incrementar la confiabilidad del estándar. Una desventaja es que consume mucho tiempo y es más complejo que otros métodos de jueceo.

Método de Livingston y Zieky (1982). Corresponde a la categoría de los métodos Empírico-Jueceo. Es también conocido como *Método de contraste de*

grupos. En lugar de enfocarse en el examen, se basa primordialmente en los examinandos. Aquí se clasifica a los examinandos en uno de dos grupos: aptos o no aptos. Después de clasificar a los examinandos, se combinan las distribuciones de las puntuaciones de ambos grupos y el punto en donde se intersectan ambas distribuciones, es el punto de corte. Este método es aparentemente sencillo y simple, pero el punto de corte puede verse afectado por diversos factores, tales como la proporción de los no calificados en la muestra, la severidad con la que los jueces definen el dominio, la decisión que se toma para clasificar los casos dudosos (Por ejemplo eliminarlos, considerarlos aptos, o considerarlos no aptos). En la tabla 2 se presenta una síntesis de los métodos descritos para establecer puntos de corte para las pruebas criterioles.

Tabla 2. Modelos continuos más comunes para establecer puntos de corte.

Procedimiento	Modelo	Características	Ventajas/Desventajas
Jueceo	Nedelsky (1954)	▪ Utilizado solo con pruebas de opción múltiple. ▪ Basado en el supuesto de que el examinando elimina en primer término las respuestas que a su juicio son incorrectas, y procede posteriormente a escoger la que considera correcta entre las que restan. Forma de obtención del PC ▪ Los jueces establecen qué opciones del ítem debería considerar erróneas el estudiante con competencia mínima; ▪ Asignan el recíproco del número de opciones no seleccionadas por el examinando; p. ej: en un ítem de 5 opciones los jueces consideran que un alumno mínimamente competente debería saber que son falsas 3 de las 5, el valor que se busca sería 0.5 (recíproco de 2). ▪ Se obtiene la suma de todos los recíprocos. ▪ El valor resultante será la puntuación del alumno mínimamente competente para el juez en cuestión. ▪ El punto de corte es la media de los valores resultantes producidos por los jueces.	Desventajas: ▪ No se puede saber con certeza si el examinando utiliza el método de adivinación o no. ▪ Los valores asignados por los jueces regularmente son .5, lo cual produce un punto de corte más bajo que otros métodos. ▪ Requiere mayor entrenamiento de los expertos, provocando que parezca ser más científico que otros métodos
	Angoff (1971)	▪ El grupo de especialistas revisa los resultados y estima la probabilidad de que un examinando pueda contestar correctamente la pregunta. Forma de obtención del PC: ▪ Se suman las estimaciones de los especialistas ▪ Se determina el promedio de las estimaciones ▪ Se utiliza el promedio para establecer el punto de corte.	Ventajas: ▪ Fácil de entender, implementar y calcular. ▪ El punto de corte es específico a las preguntas de la prueba y juicios de valor de los especialistas ▪ Método recomendado cuando se persiguen fines de certificación porque permite tomar decisiones en una línea continua de probabilidades, lo cual produce puntos de corte más directos y verosímiles. Desventajas: ▪ Generalmente se obtienen variaciones grandes entre las probabilidades de los reactivos de cada juez
	Ebel (1972)	▪ Se hace un análisis de la dificultad y relevancia de cada reactivo. Forma de obtención del PC: ▪ Implica un procedimiento de dos etapas ▪ Se estima el porcentaje de reactivos de cada celda que puede contestar correctamente un examinando. ▪ Para obtener el promedio se multiplica el número de reactivos de las celdas por el porcentaje respectivo. ▪ Se suman todos los resultados y divide entre el número de reactivos (promedio ponderado) 1.- Cada juez clasifica las preguntas en grupos, basándose en su dificultad (fácil, mediano y difícil), y su relevancia (esencial, importante, aceptable y objetable) 2.- Emiten juicios por cada uno de los 12 bloques de la tabla de clasificación, sobre el rendimiento de un examinando con competencia mínima.	Ventajas: ▪ Es un método muy valioso cuando las preguntas del examen difieren ampliamente en importancia. Desventajas: ▪ Es un método inadecuado para exámenes cortos ▪ Su procedimiento implica mucho tiempo.
Jueceo-Empírico	Jaeger (1978)	▪ Clasificado como interactivo porque los jueces que participan tienen diversas características (maestros, administradores, etc.), y porque utiliza información normativa del desempeño de los examinandos. ▪ No busca competencias mínimas, busca si el examinando puede o no contestar correctamente el reactivo ▪ Así se determina si se le otorga o no el documento que acredite su conocimiento. Forma de obtención del PC: ▪ Es el cálculo del promedio de las respuestas de los jueces.	Ventajas: ▪ El proceso interactivo de este método permite reducir la variabilidad en la distribución de los estándares recomendados, y con ello incrementar la confiabilidad del estándar. Desventajas: ▪ Consume mucho tiempo y es más complejo que otros métodos de jueceo.
Empírico-Jueceo	Livingston y Zieky (1982)	▪ También conocido como Método de contraste de grupos. ▪ Se basa primordialmente en los examinandos y no en el examen. Se clasifica a los examinandos en uno de dos grupos: aptos o no aptos. Forma de obtención del PC: ▪ Se combinan las distribuciones de las puntuaciones de ambos grupos ▪ El punto en donde se intersectan ambas distribuciones es el punto de corte.	Ventajas: ▪ Es un método sencillo y simple Desventajas: ▪ El punto de corte puede verse afectado por factores como ○ La proporción de los no calificados en la muestra, ○ La severidad con la que los jueces definen el dominio; ○ La decisión que se toma para clasificar los casos dudosos (eliminarlos, aptos, no aptos).

Capítulo III.- ESQUEMA ANALÍTICO – INFORMÁTICO PARA ELABORAR INFORMES DE RESULTADOS

Para efectuar el análisis de los datos derivados de la aplicación a gran escala del examen criterial y poder elaborar los correspondientes informes de resultados, se desarrolló un esquema (ver figura 1) conformado por elementos, procesos y relaciones que se describen en las siguientes secciones.

3. 1. Análisis curricular

Se trata de un análisis de contenido de fuentes formales e informales que definen el currículum, tales como los planes y programas de estudio, materiales instruccionales, prácticas educativas y otros aspectos que se consideran como representativos del universo de conocimiento que se evalúa. Este análisis se puede llevar a cabo de acuerdo al tipo de organización curricular de que se trate; por ejemplo, por unidad-tema-subtema, por eje – subeje – línea de formación, o por cualquier otro tipo de organización que presente el contenido del currículum.

Este análisis es de suma importancia, ya que la calidad de los reactivos de un instrumento de evaluación de tipo criterial es constantemente juzgada en relación a los resultados que pretende el currículum. El proceso de análisis curricular culmina con un producto notable llamado retícula, que en forma sintética y gráfica permite identificar y mostrar la estructura del contenido que evalúa el examen por niveles

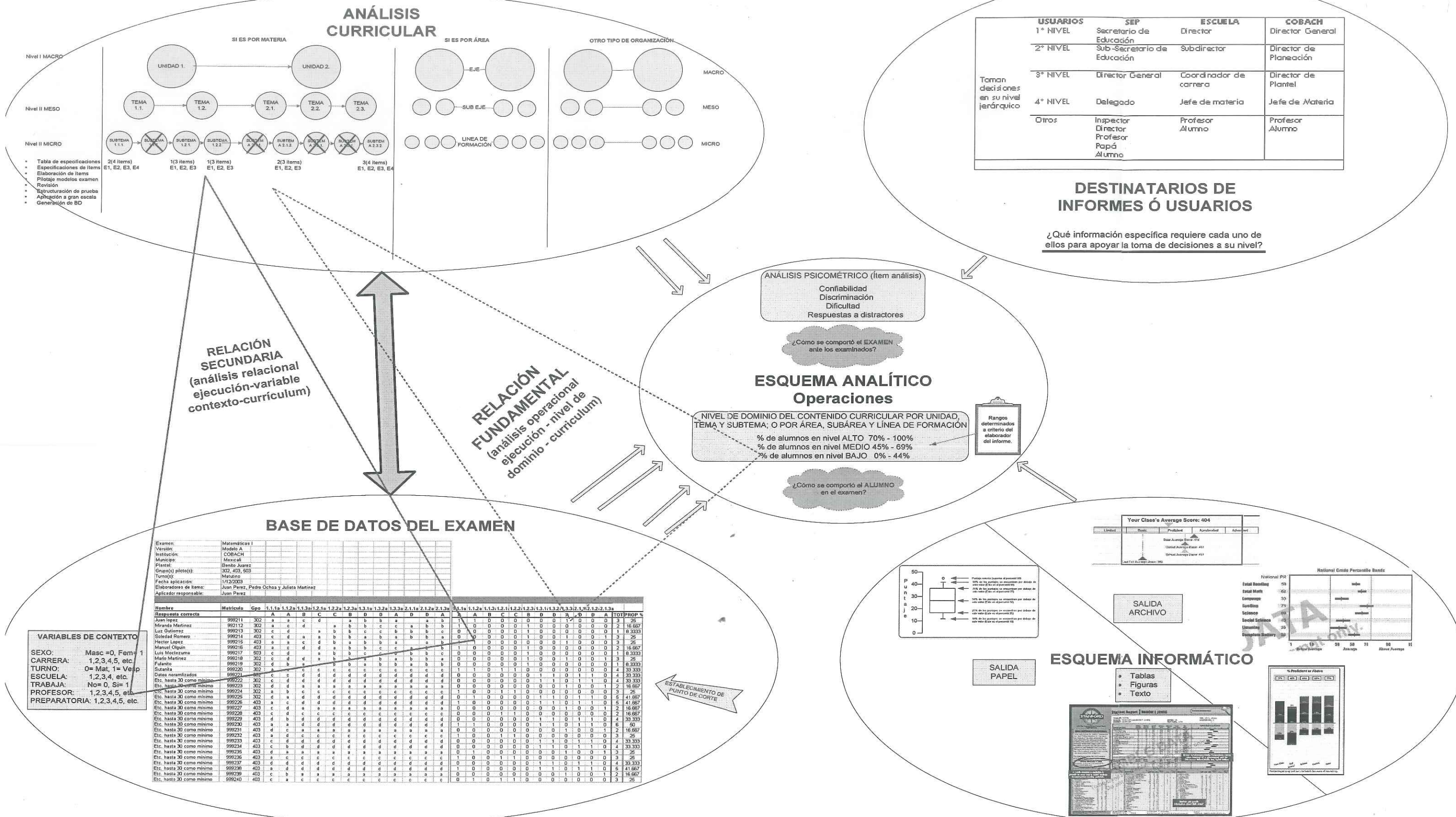


Figura 1. Esquema analítico - informático de los resultados de la aplicación a gran escala de un examen criterial

denominados macro, meso y micro. En la parte superior izquierda de la figura 1, se ilustra un ejemplo bosquejado de una retícula. Otros productos que se derivan del análisis curricular, son: la tabla de especificaciones del examen, las especificaciones de cada uno de los reactivos que conforman la prueba, los reactivos elaborados a partir de dichas especificaciones y la estructuración de los ítems en los diversos modelos de la prueba que servirán para evaluar el dominio que tienen los examinandos del contenido curricular. De esta manera se cuenta con un análisis detallado del contenido importante a evaluar, así como los documentos que justifican dicho análisis.

3. 2. Base de datos del examen

Constituida por una tabla de doble entrada que contiene varios elementos, entre los cuales se encuentran los siguientes:

- **EXAMINANDOS.** Se especifican los nombres o forma de identificar a los examinandos en los renglones de la tabla.
- **ÍTEMS.** Se presentan en las columnas de la tabla.
- **DATOS CRUDOS.** En consecuencia, cada celda contendrá la letra que corresponde a la opción (a, b, c, d, etc.; según sean tres, cuatro o más opciones de respuesta) que seleccionó cada alumno en cada reactivo de opción múltiple del examen. Éste se considera el dato primario o crudo de la prueba.

- CALIFICACIÓN DE LA PRUEBA Y OBTENCIÓN DEL PUNTAJE TOTAL.

Posteriormente, estos datos se convertirán en datos interpretables mediante un procedimiento que inicia con la calificación de la prueba, la que a su vez comienza con la conversión de las letras en las celdas (a, b, c, d) en ceros y unos, mediante la comparación de las respuestas de cada examinando con la clave de respuestas correctas; siendo para cada examinando, en cada reactivo: “cero” incorrecto y “uno” correcto. Enseguida, al final de la base de datos se presenta una columna con puntajes totales logrados por los examinandos, mismos que se obtienen mediante la sumatoria de los aciertos conseguidos en los ítems por cada alumno. Para ser interpretables, estos datos se deberán convertir en algún tipo de puntuación referida a un criterio. En las pruebas criterioles, el dato “duro” que se utiliza comúnmente para informar sobre los resultados de la prueba es el porcentaje de aciertos, debido a que brinda una síntesis de la ejecución en el examen que resulta familiar y permite ubicar rápidamente el nivel de dominio logrado en la prueba. Usualmente, también se presentan los porcentajes de aciertos en rangos que forman una escala ordinal, a fin de representar de manera cualitativa estándares de ejecución o niveles de dominio (por ejemplo, alto, medio y bajo) de los contenidos que miden los ítems de la prueba.

3. 3. Variables de contexto

De manera relacionada, el porcentaje de aciertos en cada área del currículum puede ser vinculado con variables contextuales del propio examinando, de la escuela, de su entorno u hogar, de las condiciones de operación del currículum y otras más, a fin de que la actuación del examinando resulte más significativa.

Para efectos de análisis estadístico de los datos, comúnmente se lleva a cabo una codificación de dichas variables, mediante la conversión de los datos a una escala numérica, según sea el caso. Por ejemplo, en la variable de género, se puede codificar femenino con un “cero” y masculino con “uno”; la variable de carrera con uno, dos, tres, cuatro, etc., según la diversidad de carreras que cursen los examinandos.

Así, este análisis relacional pretende descubrir las posibles relaciones entre las variables de contexto, con la ejecución del examinando en el examen, referida al dominio del currículum. En la figura 1 dichos vínculos se representan mediante un triángulo de relaciones entre los resultados en la base de datos, el análisis curricular y las variables de contexto. Para llevar a cabo este tipo de análisis es necesario asociar variables de contexto con los datos que se obtienen de la ejecución de los examinandos, de tal manera que los resultados se tornen más significativos. En este sentido, es importante considerar la relación existente entre las respuestas de los examinandos y variables tales como tipo de escuela, nivel socioeconómico del

alumno, nivel escolar de los padres, género, turno, maestro que imparte la materia, si trabaja o no, etc.

A continuación, se presentan algunas variables contextuales que se clasificaron como producto de la literatura revisada, y que usualmente se relacionan con las puntuaciones obtenidas por los examinandos.

- **Variables propias del alumno:** entre este tipo de variables se encuentran el género, la edad, la escuela de procedencia, si trabaja o no trabaja, el estado civil, etc.
- **Variables propias de la Escuela:** incluyen el turno, la carrera que estudia el examinando, la zona escolar a la que pertenece la escuela, el nivel educativo que se imparte en el plantel, el tipo de control o sistema al que pertenece la escuela (pública, privada, abierta, etc.), y otros.
- **Variables propias del hogar en el que vive el examinando:** este tipo de variables incluyen aspectos del entorno del examinando, tales como la escolaridad de la madre, ocupación del padre, ingreso familiar mensual, cantidad de miembros en la familia, la zona en la que vive, el tipo de vivienda, etc.
- **Variables propias de la operación del currículum:** aquí se pueden incluir el estilo didáctico de quien imparte la materia, si cuenta o no con tutor académico, la disponibilidad de libros de texto y el perfil docente, entre otros.

Cualquier variable perteneciente a alguno de los tipos mencionados en los párrafos anteriores puede ser asociada con la puntuación obtenida en la prueba. Sin embargo, como se mencionó anteriormente, debe efectuarse una codificación de cada variable pues el análisis estadístico correlacional requiere que ambas variables que se relacionan se encuentren en una escala. Cuando ello sucede, es posible establecer una medida de correlación entre la variable de contexto de interés y la puntuación obtenida por los examinandos en la prueba, así como también el nivel de significancia estadística de dicha relación.

Para realizar este análisis correlacional puede utilizarse una variedad de programas de análisis estadístico, tales como SPSS, EZAnalyze (de distribución libre en <http://EZAnalyze.com>), o Excel. Aunque es muy común utilizar el programa SPSS para realizar una gran variedad de análisis estadísticos, su uso es relativamente complejo y especializado. Por ello, para el caso de aquellas personas que se encuentran dentro del área educativa y que no tienen muchos antecedentes en estadística, se recomienda el uso del programa Excel debido a que es accesible y su manejo resulta relativamente intuitivo, comparado con otros programas de cómputo más sofisticados.

Al utilizar el programa Excel para realizar el mencionado análisis de correlaciones, se requiere agregar un componente que usualmente no se encuentra instalado en las versiones básicas del programa. Dicho componente se denomina Análisis de Datos (Data Analysis) y se importa con relativa facilidad desde el menú

de Herramientas (Tools); enseguida se selecciona la opción Complementos (Add-Ins) y a continuación se selecciona la opción Paquete de Herramientas de Análisis (Analysis ToolPack). De este modo, en el menú de Herramientas aparecerá disponible la opción Análisis de Datos (Data Analysis), misma que puede ser utilizada para estimar el índice de correlación entre la variable contextual analizada y el puntaje de los examinandos en la prueba, y que también indica el nivel de significancia estadística que tiene dicha correlación, lo cual aporta un elemento muy importante para la interpretación de los resultados. Es importante mencionar que para obtener ambas estimaciones es necesario utilizar la opción Regresión (Regression) tras seleccionar Análisis de Datos (Data Analysis), ya que si se utiliza la función Correl incluida en la opción Función (Function) del menú Insertar (Insert), el programa solamente proporcionará la correlación entre las variables de interés, pero no incluirá el dato del nivel de significancia.

Ahora bien, la mayoría de los informes de resultados incluyen también otro tipo de información contextual respecto a variables existentes en la escuela o en el programa educativo, lo cual sirve para proveer un contexto más amplio para ayudar a la interpretación de resultados.

Estos datos contextuales pueden ser de utilidad para que el lector lleve a cabo una evaluación mucho más acertada de la ejecución del programa educativo en cuestión. Los informes de resultados generalmente incluyen información de variables como cuestiones demográficas de los estudiantes, datos socioeconómicos de la

comunidad que los rodea, y algunos aspectos económicos relacionados con dicha comunidad; sin embargo, una aproximación más directa que apoya la interpretación de los resultados es la inclusión de un breve texto que contenga una explicación narrativa de lo que se encuentra en el informe de resultados.

Cabe destacar que, aunque la información contextual y descriptiva son elementos importantes de los informes de resultados, su función es meramente de apoyo, y deberá cuidarse que no opaquen en ningún momento los datos sobre los resultados de la evaluación, los cuales siempre deberán ser el foco central de un informe de resultados. En este sentido, es importante mencionar que las investigaciones han demostrado que la mayoría de los destinatarios (incluyendo el público en general), buscan que el enfoque central de los informes sea precisamente los resultados; por ello si la información descriptiva es excesiva, la comprensión de los informes de resultados se tornará difícil (Condelli, 2005).

3. 4. Análisis de la ejecución de los examinandos

En un examen criterial alineado con el currículum queremos dar cuenta de en qué medida los resultados obtenidos por los examinandos reflejan el dominio de dicho currículum. En consecuencia, después de obtener el porcentaje total de aciertos, es necesario determinar el porcentaje de aciertos que tuvo cada examinando en los grupos de ítems que evaluaron el dominio de cada unidad y tema (o de cada categoría en que esté organizado el contenido, según la estructura del

currículum). Lo anterior con la finalidad de saber qué segmentos del currículum domina el alumno, y qué segmentos no domina. Esta relación es de primera importancia, ya que muestra el vínculo estrecho que guardan el análisis curricular y la base de datos.

En la figura 1 dicha relación se representa mediante un triángulo de relaciones fundamentales que se denomina análisis operacional, pues considera las operaciones necesarias para evaluar el nivel de ejecución de los examinandos en la prueba, referido al dominio del currículum, desagregado por área que evalúa la prueba. Este análisis considera también las evidencias recabadas sobre la calidad técnica del instrumento, mismas que pondrán un límite a la confianza que podemos tener en los resultados de la prueba y a la validez de las inferencias que podremos llevar a cabo a partir de ellos.

3. 5. Destinatarios de informes de resultados.

Los destinatarios de los informes de resultados son un elemento clave que sirve para determinar el nivel de profundidad y detalle con el que se elaborarán tanto los análisis de datos como el informe, así como los tipos de información que deberán incluir. Es necesario hacer una descripción detallada de las características de cada uno de los destinatarios a los cuales se dirige el informe, del tipo de información que necesitan y de los usos que le darán a la información.

Lo anterior significa que una vez que se tiene el análisis de la ejecución de los examinandos, se deben elaborar informes de resultados para los usuarios, quienes toman decisiones en su nivel jerárquico, y por ende tienen necesidades informativas diferenciales para fundamentar la toma de dichas decisiones. En consecuencia, diversos modos de organización institucional nos proveen diferentes modelos de destinatarios.

Por ejemplo, para un funcionario de primer nivel en una institución educativa se deberá elaborar un informe sobre la ejecución de la población general que fue examinada, clasificada por municipio, carrera y unidades temáticas. Se podrá relacionar la ejecución general de los examinandos con algunas variables de contexto, tales como tipo de escuela, sexo, turno, etc. Este informe no debe ser muy detallado, pues el usuario toma decisiones generales respecto a la planeación de la institución.

A funcionarios de niveles jerárquicos inferiores (coordinadores de carrera, jefes de materia, docentes, etc.), se les proporcionará información más específica; por ejemplo, porcentajes de aciertos de los alumnos de la carrera o materia; por área curricular evaluada y de los alumnos de todos los maestros que imparten la materia, con la finalidad de que se puedan comparar y mejorar estrategias docentes.

Como producto de la literatura revisada, y con el propósito de simplificar la clasificación de los destinatarios, se proponen cuatro niveles:

Primer nivel: si la evaluación es de tipo Estatal, el informe deberá dirigirse al Secretario de Educación del Estado, o al Director General de la institución. En estos casos, tanto los análisis como el informe deben incluir información respecto a:

- Distribución de promedios de las escuelas del Estado, en cada área del currículum evaluada.
- Distribución de promedios de las escuelas, por municipio, en cada área curricular.
- Tendencia del municipio y promedios estatales en cada área curricular, durante los cinco años anteriores.
- Gráficas que comparen simultáneamente los resultados de logro contra acciones educativas determinadas (perfil de profesores, recursos de la escuela, etc.)

Si la evaluación es de tipo departamental o colegiada, el informe deberá dirigirse al Director del plantel, en cuyo caso deberá incluir el mismo tipo de información, pero haciendo referencias al nivel de la escuela de que se trate.

Segundo nivel: si la evaluación es de tipo Estatal, departamental o colegiada, el informe deberá dirigirse al Sub-Secretario de Educación del Estado, al Director general de planeación educativa ó al Subdirector del plantel; en cuyo caso deberá incluir información respecto a:

- Promedio de ejecución de estudiantes en grupos de preguntas que evalúan blancos de aprendizaje claves, por área del currículum.

- Promedio de ejecución de estudiantes en cada área del currículum.
- Resultados de ambos aspectos, por municipio o, en su caso por carrera.

Tercer nivel: si la evaluación es de tipo Estatal, departamental o colegiada, el informe deberá dirigirse al Director General, Delegado educativo, Inspector escolar, Coordinador de carrera, ó Director de plantel, y deberá incluir información acerca de:

- Promedio de resultados por escuela dentro de la región de servicio, para cada área del currículum
- Promedio de resultados por grupo, en cada unidad y tema
- Tendencia de la ejecución promedio, durante los cinco años anteriores

Cuarto nivel: en este nivel se incluyen al Jefe de materia, profesores, y otros, según el tipo de evaluación. Este informe deberá incluir datos respecto a:

- Promedio general y por género, de resultados de los estudiantes de la escuela, por área del currículum evaluada.
- Porcentaje de aciertos por área del currículum, para cada examinando.
- Comparación del promedio general y por género, de los resultados de los estudiantes de la escuela, por área del currículum, contra las correspondientes medias de las escuelas del Estado.

Se considera que pueden existir otros niveles, en los cuales se encuentran destinatarios tales como los padres de familia, alumnos y público en general. Para este tipo de informes el nivel de especificidad y detalle de la información deberá ser adaptada para las características específicas y necesidades informativas que tengan.

Existen diversos procedimientos que se pueden utilizar cuando se elaboran informes de resultados de un examen de gran escala, y la elección de ellos depende críticamente del tipo de destinatario al cual está dirigida la información. Por su parte, el aspecto más importante sobre el cual versará el informe es el tipo de información que resulta más relevante para apoyar la toma de decisiones en el nivel correspondiente a cada destinatario. A continuación se presentan los principales usuarios de la información del examen y los principales datos que son de su incumbencia:

- **Estudiantes:** al reportarles sus resultados se recomienda hacerlo en dos fases: la primera implica proporcionar un informe general de la ejecución de los examinandos que ya recibieron sus resultados individuales. La segunda se trata de sesiones de seguimiento individual con cada estudiante, enfocadas sobre la forma en que los maestros abordarán las necesidades individuales de los estudiantes.
- **Padres de familia:** dado el interés para los padres de familia por conocer el rendimiento académico de sus hijos, tanto en forma individual como en relación a todo el estudiantado y otras escuelas, los informes deben incluir estos aspectos como punto medular. Para ello se sugieren cuatro estrategias:
 1. Reuniones individuales entre el maestro y los padres de familia. No obstante, aunque es lo más recomendable, debido a situaciones particulares no siempre es posible llevarlas a cabo.

2. Informes individuales enviados a casa. Pueden proporcionar una forma de comunicación accesible maestro – padre de familia, pero deberán incluir qué tan buena o mala fue la ejecución del examinando, así como los pasos a seguir después de la evaluación, que incluyen la forma en que los padres pueden participar en un plan de acción para apoyar la educación de sus hijos.
 3. Reuniones con padres de familia.
 4. Boletines informativos dirigidos específicamente a los padres de familia.
- **Funcionario de primer nivel (Secretario de Educación, Rector, Director General, etc.)** Debido a que se trata de entidades que se encargan del establecimiento de políticas educativas a nivel estatal o institucional, el tipo de información reportada debe incluir en primer lugar, una distribución de promedios o porcentajes de aciertos de los planteles del estado, municipio, campus, región, institución, etc., por cada área curricular. En segundo término, se debe reportar la tendencia del municipio, campus, región, etc., y los promedios en cada eje curricular, durante los cinco años anteriores. Por último, deberá incluirse en forma gráfica, una comparación simultánea de los resultados de logro vs. las acciones educativas determinadas (perfil de profesores, recursos de la escuela, etc.). Para ello se sugiere una estrategia que implica tres fases:

En la primera fase se debe elaborar un informe de los antecedentes que rodean la evaluación, que incluya una explicación de lo que fue evaluado, los tipos

de evaluación empleados y motivos por los cuales se eligieron, así como la forma en que se utilizarán y reportarán los resultados. Este informe deberá ser entregado en el momento en el que se esté recolectando la información, y antes de que los resultados de la evaluación estén disponibles. De esta forma, se invita a este tipo de destinatario a enfocarse en el trasfondo de la evaluación, y no únicamente en los datos cuantitativos.

En la segunda fase, se incluirá un informe que contenga los resultados de la evaluación nacional, estatal, municipal, etc., y que responda a preguntas tales como:

- ¿Qué resultados se obtuvieron respecto a la ejecución a nivel nacional, estatal, municipal, etc.?
- ¿Cómo se presentan los resultados en comparación al año anterior?
- ¿En qué nivel se encuentran los puntajes de los examinandos: básico, medio, alto?
- ¿Qué medidas se están tomando para examinar los resultados de la evaluación y mejorar el proceso educativo?
- ¿Cómo se elaborarán los informes que serán dirigidos a los padres de familia y público en general?

El tercer informe debe proporcionar un seguimiento de los esfuerzos realizados para mejorar el proceso educativo en el estado o, en su caso, el proceso instruccional en la escuela, así como de su efectividad. Aunque se propone como un informe opcional, es muy efectivo para orientar a este tipo de destinatario hacia el

verdadero propósito de la evaluación, el cual es mejorar el proceso educativo, y no solamente proporcionar puntajes para establecer niveles de calidad en las escuelas.

- **Directores de planeación educativa, directores académicos y otros similares.**

Los datos más relevantes incluyen: promedios de ejecución de estudiantes en cada eje curricular y en grupos de preguntas que evalúan blancos de aprendizaje claves por eje curricular; resultados en ambos por campus, región, municipio, etc.

- **Inspector, coordinador, etc.** La información que resulta relevante en este nivel incluye los promedios de resultados por plantel dentro del ámbito o región de servicio, por cada eje curricular; así como la tendencia de la ejecución promedio, durante los cinco años anteriores.

- **Director de plantel.** El tipo de información relevante que debe reportarse es más específico, ya que lo que interesa es el comportamiento de los examinandos dentro de una escuela específica. Incluye resultados de los examinandos del plantel en cuanto a promedio general y por género; porcentaje de aciertos de cada examinando; comparación del promedio general y por género, de los resultados de los examinandos del plantel, contra las correspondientes medias de los planteles del campus, región, municipio, etc. Todo lo anterior por eje, subeje curricular ó línea de formación que define el currículum (Contreras y Backhoff, 2004). Debido a que este tipo de destinatario tiene relación directa con la implementación de estrategias y acciones encaminadas al mejoramiento académico del alumno, es importante incluir en el informe elementos que proporcionen directrices para llevar

a cabo sesiones de seguimiento individual con cada estudiante, enfocadas en la forma en que los maestros abordarán las necesidades individuales de los mismos.

- **Público en general:** lo primero que debe hacerse es decidir el tipo de público al cual se va a dirigir la información, ya que es un grupo muy heterogéneo que tiene distintas necesidades de información y conocimiento. Enseguida se deben explicar las metas y propósitos del informe. Por último, se deben establecer los medios por los cuales se dará a conocer la información.

La tabla 3 resume la información de los principales destinatarios de los informes de resultados del examen aquí presentada.

3. 5. 1. Análisis, reporte e interpretación de resultados en forma desagregada

Actualmente se utilizan los resultados de las pruebas criterioles en áreas muy diversas. Por ejemplo, los maestros, utilizan los resultados de la ejecución de los alumnos para ubicarlos en programas educativos adecuados a sus necesidades, monitorear sus avances, realimentar su aprendizaje, identificar sus deficiencias y también para planear su instrucción. En el área de educación especial este tipo de informe ayuda mucho en el diagnóstico de problemas de aprendizaje, así como en el monitoreo del progreso de los alumnos. También se utiliza este tipo de informes para evaluar programas educativos normales y avanzados. Pero quizás uno de los usos más importantes que se le da a los resultados de las pruebas se ubica en el área de certificación profesional (Hambleton, 1988).

Tabla 3. Destinatarios de los informes de resultados del examen

Tipo de organización	Tipo de examen	Destinatario	Tipo de información a ser reportada
Secretaría estatal de educación	Examen estatal	Primer nivel: Secretario de Educación del Estado y Subsecretario de Educación Básica	1. Distribución de promedios de las escuelas del estado, en cada eje curricular 2. Distribución de promedios de las escuelas, por municipio, en cada eje curricular 3. Tendencia del municipio y promedios estatales en cada eje curricular, durante los cinco años anteriores 4. Gráfica que compara simultáneamente los resultados de logro contra acciones educativas determinadas (perfil de profesores, recursos de la escuela, etc.)
		Segundo nivel: Director general de planeación educativa	1. Promedio de ejecución de estudiantes en grupos de preguntas que evalúan blancos de aprendizaje claves, por eje curricular 2. Promedio de ejecución de estudiantes en cada eje curricular 3. Resultados en 1 y 2, por municipio
		Tercer nivel: Inspector	1. Promedio de resultados por escuela dentro de la región de servicio, para cada eje curricular 2. Tendencia de la ejecución promedio, durante los cinco años anteriores
		Cuarto nivel: Director de Escuela, subdirector y profesor	1. Promedio general y por género, de resultados de los estudiantes de la escuela, por eje curricular 2. Porcentaje de aciertos por eje curricular, para cada examinando 3. Comparación del promedio general y por género, de los resultados de los estudiantes de la escuela, por eje curricular, contra las correspondientes medias de las escuelas del estado
		Otros niveles: Estudiante y padre	1. Porcentaje de aciertos por eje curricular, obtenidos por el estudiante 2. Comparación del porcentaje promedio de aciertos del grupo, por eje curricular, contra el correspondiente logrado por el estudiante
Institución educativa estatal	Examen estatal	Primer nivel: Director general	1. Distribución de promedios de las escuelas del estado, en cada unidad 2. Distribución de promedios de las escuelas, por municipio, en cada unidad 3. Tendencia del municipio y promedios estatales en cada unidad, durante los cinco años anteriores 4. Gráfica que compara simultáneamente los resultados de logro contra acciones educativas determinadas (perfil de profesores, recursos de la escuela, etc.) 5. Porcentaje de aciertos de las escuelas del estado, por municipio, en cada unidad
		Segundo nivel: Director de planeación académica	1. Promedio de ejecución de estudiantes en grupos de preguntas que evalúan blancos de aprendizaje claves, por unidad y tema 2. Promedio de ejecución de estudiantes en cada unidad y tema 3. Resultados en 1 y 2, por municipio 4. Porcentaje de aciertos de los estudiantes, en cada unidad y tema evaluados
		Tercer nivel: Director de plantel	1. Promedio de resultados por grupo, para cada unidad y tema 2. Tendencia de la ejecución promedio, durante los cinco años anteriores
		Cuarto nivel: Jefe de materia y profesor	1. Promedio general y por género, de resultados de los estudiantes de la escuela, por unidad y tema 2. Porcentaje de aciertos por unidad y tema, para cada examinando 3. Comparación del promedio general y por género, de los resultados de los estudiantes de la escuela, por unidad y tema, contra las correspondientes medias de las escuelas del estado
		Otros niveles: Estudiante	1. Porcentaje de aciertos por unidad y tema, obtenidos por el estudiante 2. Comparación del porcentaje promedio de aciertos del grupo, por unidad y tema, contra los correspondientes logrados por el estudiante
Escuela	Examen departamental o colegiado	Primer nivel: Director	1. Distribución de promedios de las carreras de la escuela, en cada unidad 2. Tendencia de la escuela y promedios en cada carrera, durante los cinco años anteriores 3. Gráfica que compara simultáneamente los resultados de logro contra acciones educativas determinadas (perfil de profesores, recursos de la escuela, etc.)
		Segundo nivel: Subdirector	1. Promedio de ejecución de estudiantes en grupos de preguntas que evalúan blancos de aprendizaje claves, por unidad y tema 2. Promedio de ejecución de estudiantes en cada unidad y tema 3. Resultados en 1 y 2, por carrera
		Tercer nivel: Coordinador de carrera	1. Promedio de resultados por grupo, para cada unidad y tema 2. Tendencia de la ejecución promedio, durante los cinco años anteriores
		Cuarto nivel: Jefe de materia y profesor	1. Promedio general y por género, de resultados de los estudiantes de la escuela, por unidad y tema 2. Porcentaje de aciertos por unidad y tema, para cada examinando 3. Comparación del promedio general y por género, de los resultados de los estudiantes de la carrera, por unidad y tema, contra las correspondientes medias de las carreras de la escuela
		Otros niveles: Estudiante	1. Porcentaje de aciertos por unidad y tema, obtenidos por el estudiante 2. Comparación del porcentaje promedio de aciertos del grupo, por unidad y tema, contra los correspondientes logrados por el estudiante

En el ámbito internacional, en 1994 el Congreso de los Estados Unidos aprobó el Acta de las Escuelas de América documento en el cual se estipula en el Título I, como una obligación, y a la vez como una oportunidad para los estados, organismos educativos y escuelas, desagregar y analizar los datos resultantes de las evaluaciones de sus estudiantes (Jaeger y Tucker, 1998). Estas evaluaciones permitirán la desagregación de la información en variables específicas, tales como género, escuela, desempeño en el área de matemáticas, estatus migratorio, discapacidades, situación económica, etc. Sin embargo, esto implica una necesidad de definir adecuadamente dichas variables. Entre otras cosas, es necesario especificar definiciones de las categorías de los estudiantes, de los aspectos legales en cuanto a la confidencialidad de los resultados, de la generalización adecuada de los resultados de la evaluación, de la interpretación válida de los resultados de la evaluación dentro y entre las categorías de los estudiantes, y de la difusión de los resultados de tal forma que sean útiles a distintos destinatarios.

Lo anterior brinda una gran oportunidad para llevar a cabo un análisis nuevo y creativo de la información, cuyo uso podría extenderse al análisis de las políticas educativas actuales, evaluación de los programas existentes, planeación y desarrollo de intervenciones instruccionales innovadoras, así como la elaboración de informes respecto al estado actual y al progreso de los alumnos. Quienes formen parte activa del sistema educativo y tengan injerencia en el análisis y reporte del desempeño colectivo de los alumnos, se verán beneficiados al desagregar la información que

arrojan los resultados de las evaluaciones. Dentro de esa población se encuentran maestros, directivos, inspectores de zona, administradores, instituciones de educación y personal del área de la evaluación educativa, e inclusive agencias gubernamentales que toman decisiones en torno a políticas educativas a nivel local, estatal y/o nacional (Jaeger y Tucker, 1998).

Jaeger y Tucker (1998) proponen tres escenarios para elaborar informes de resultados: la Zona escolar, la Escuela y el Estado. Los autores explican cuáles son los elementos y procedimientos que se deben incluir en todo informe, y la importancia de reportar en forma desagregada dicha información.

Zona escolar

En el caso de la zona escolar, es necesario abordar los siguientes aspectos:

1. La forma en que se desagregarán los resultados de la prueba para entender de mejor manera los factores asociados al desempeño de los estudiantes.
2. La manera en que se presentarán dichos resultados, de tal forma que se resalten los aspectos más interesantes de la ejecución de los alumnos.
3. Si el tamaño del grupo de estudiantes evaluados es lo suficientemente grande como para respaldar el informe.
4. Si las diferencias entre las ejecuciones promedio de los estudiantes de varios grupos es lo suficientemente confiable.

5. El tipo de conclusiones que se pueden obtener respecto a las causas de los resultados obtenidos.

De esta manera se podrán obtener informes de resultados cuyas principales funciones sean proporcionar información útil para que los directores de planteles y maestros de una zona escolar específica puedan llevar a cabo una adecuada planeación curricular, para que las autoridades educativas utilicen la información cuando se realicen revisiones de las políticas educativas, y también para que se difundan los resultados entre la población que conforma dicha zona escolar, así como entre los padres de familia cuyos hijos se encuentran estudiando dentro de ella.

Para que esto pueda llevarse a cabo, se requiere analizar los resultados de cada zona escolar que se encuentra en el municipio y en el estado, y los de cada escuela en cada zona, en forma separada.

Se debe examinar a los alumnos por lo menos una vez en el nivel escolar que se esté evaluando (básico, medio, medio superior o superior); y los resultados de la evaluación se deben analizar y reportar por zona y escuela en forma separada, ubicando por categorías a los estudiantes (género, nivel socioeconómico, turno, etc.). Para que esta información sea útil para los maestros cuando lleven a cabo su planeación curricular, se sugiere que:

1. Se resuman los resultados de la evaluación de los alumnos de cada grupo.

2. Una parte del resumen describa la ejecución general del grupo en lo que respecta a la materia evaluada.
3. Otro tipo de información que debe incluir el informe es si la ejecución de los alumnos fue similar, o si algunos alumnos tuvieron mejores resultados que otros.
4. Es muy importante que al reportar los resultados se especifique la ejecución de los alumnos en relación a los reactivos que evalúan dominios específicos del conocimiento general que evalúa la prueba.
5. También se debe incluir un marco de referencia para que el docente sepa discriminar entre los puntajes que distinguen una buena ejecución de una excelente; y, si se encuentra disponible, establecer una comparación entre puntajes actuales con puntajes de un año anterior.

En el caso de los Directores de plantel, para que la información de los resultados de la evaluación resulte útil en la planeación curricular, se sugiere que:

1. Se genere un documento que sintetice los resultados de los estudiantes en los momentos académicos críticos, según el nivel educativo que estén cursando, de tal manera que cada Director conozca los resultados de su plantel y de los otros que conforman la zona, para que pueda establecer puntos de comparación.
2. Se disponga la información para que se pueda establecer una comparación entre la ejecución de años anteriores y el actual.

3. Al reportar la información se pueda observar si existen diferencias entre los puntajes de los examinandos, con la finalidad de saber si la ejecución fue similar para todos o no.
4. Se establezcan en el informe las categorías de estudiantes cuyo porcentaje de eficiencia se considere pobre, mediocre, bueno y excelente (o la escala correspondiente).
5. Se agrupe la información por variables de contexto, tales como género, turno, nivel socioeconómico, etc.

Existen cuatro tipos de análisis que se pueden producir con los datos: el primero es una categorización de los porcentajes obtenidos con las puntuaciones de los examinandos, agrupándolos por variable de contexto (edad, género, turno, etc.), y según la escala que se esté utilizando (p. ej. excelente, bueno, regular, malo). El segundo es la obtención de los puntajes promedio, por variable. El tercer análisis implica el uso de gráficas, donde una de las más útiles es la de cajas y bigotes (ver figura 2) que presenta la información en rangos percentiles, ya que facilita la comparación de la distribución de los puntajes de diversos grupos. Por último, se puede hacer un análisis estadístico de la variación de los puntajes promedio generada por un muestreo aleatorio y por una prueba poco confiable.

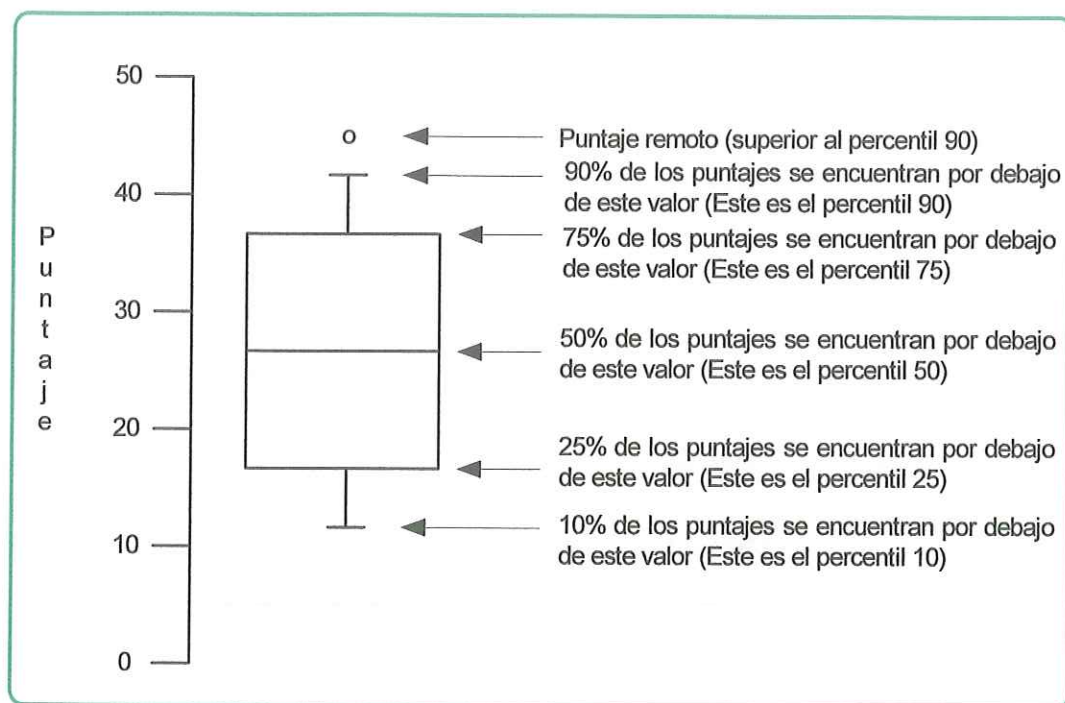


Figura 2. Ejemplo de gráfica de cajas y bigotes.

Fuente: Jaeger, R. M. y Tucker, C. G. (1998). Analyzing, disaggregating, reporting, and interpreting student's achievement test results: a guide to practice for Title I and beyond. Estados Unidos: Council of Chief State School Officers.

Escuela

En el caso del informe a una escuela, es necesario plantear una serie de preguntas que servirán como guía al realizar un análisis de datos:

1. ¿De qué manera se pueden desagregar los resultados para tener una mejor comprensión de los factores asociados con los avances colectivos de los estudiantes en un periodo de dos años?
2. ¿Cómo se puede llevar a cabo una comparación de los resultados de la ejecución de los alumnos en un periodo de dos años, de tal forma que

refleje los avances de toda la escuela en relación a las metas académicas definidas?

3. ¿Qué tan grande debe ser el grupo de alumnos como para apoyar el informe de su ejecución global en el examen?
4. ¿Qué tan confiables son las diferencias entre los promedios de la ejecución de los estudiantes en varios años consecutivos?

Para poder desarrollar una estrategia de análisis de datos, deben revisarse aspectos tales como la cantidad de alumnos evaluados en los diferentes ciclos escolares, las variables de contexto de los examinandos, y la confiabilidad de los datos obtenidos. En este sentido, es recomendable realizar el siguiente análisis estadístico:

- Distribución de frecuencias de los examinandos, por ciclo escolar y variables de contexto de las cuales se tenga información (edad, grupo, género, nivel socioeconómico, turno, etc.)
- Puntajes promedio y desviaciones estándar de todos los examinandos, de los examinandos agrupados por variables de contexto, y de cada ciclo escolar.
- Distribución de los porcentajes de estudiantes ubicados en cada nivel de dominio (previamente determinado), por ciclo escolar y por variables de contexto-ciclo escolar.
- Intervalos de confianza de los puntajes promedio globales de cada ciclo escolar evaluado, así como por variables de contexto; o una prueba de hipótesis de las

diferencias entre los puntajes promedio de los ciclos escolares que se deseen comparar.

Una vez que se lleve a cabo el análisis estadístico, se debe proceder a graficar los resultados, de tal manera que éstos se puedan apreciar en forma abreviada y clara. Se recomienda utilizar gráficas sencillas (barras, columnas, etc.). Como resultado final se pueden obtener propuestas de estrategias de mejoramiento del currículum, tales como revisiones exhaustivas de los contenidos de los cursos, materias o conocimientos que se evalúan, en contraste con el currículum propuesto, con la finalidad de saber si los estudiantes están aprendiendo lo que se supone; ó la revisión de la ejecución de los alumnos en exámenes parciales para determinar si existe la necesidad de crear programas remediales que aborden contenidos específicos en donde los alumnos estén teniendo dificultades.

Estado

El tipo de cuestionamientos que se deben abordar para realizar el análisis de datos y elaboración de informes cuando se tiene información de evaluaciones a nivel estatal, son:

1. ¿Qué tipo de análisis se pueden llevar a cabo con la finalidad de vislumbrar las tendencias de ejecución de los estudiantes en una determinada cantidad de ciclos escolares?

2. ¿De qué manera se pueden desagregar los resultados para tener una mejor comprensión de los factores asociados con los avances globales de los estudiantes en una determinada cantidad de ciclos escolares?
3. ¿Cuándo resultan estadísticamente confiables y académicamente significativos los promedios de ejecución de los estudiantes en evaluaciones de ciclos escolares consecutivos?
4. ¿Cuáles son los factores que pueden contribuir año tras año a las tendencias de la ejecución de los examinandos?
5. ¿De qué manera se puede modelar la tendencia de la ejecución global de los examinandos?

Las personas que toman decisiones y llevan a cabo políticas educativas requieren informes de resultados que reflejen tendencias de ejecución de los estudiantes a través de diferentes ciclos escolares, de tal manera que puedan determinar si existen evidencias de cambios sustanciales a través del tiempo, o evidencia de diferencias entre la ejecución evaluada y la ejecución por grupos o variables de contexto (género, turno, escuela, nivel socioeconómico, escolaridad de los padres, etc.). En pocas palabras, a esta población le interesa saber quiénes son los que cuentan con las habilidades que establecidas por el currículum, y quiénes no.

Para elaborar este tipo de informes, se debe recopilar información de evaluaciones realizadas durante varios años consecutivos, así como realizar los análisis estadísticos y gráficos necesarios. Cabe destacar que en la medida en que

se desagrega la información, se debe definir con mayor claridad a los subgrupos de los cuales se reporta la información, ya que de no hacerlo, se corre el riesgo de generar un informe, así como una interpretación de los resultados equivocados.

En conclusión, la función principal de los informes que desagregan la información recabada de una evaluación a gran escala es hacer visibles los avances (o retrasos) de diversos subgrupos de estudiantes dentro de un sistema educativo, de tal manera que brinden una oportunidad para dirigir adecuadamente los recursos con los que se cuentan dentro del campo de la educación. Dichos informes deben abordar los siguientes aspectos:

- ✓ Definiciones de las categorías para la desagregación
- ✓ Tamaño mínimo de los grupos con los cuales se reportarán los resultados
- ✓ Error de medida y de muestreo
- ✓ Construcción de los intervalos de confiabilidad
- ✓ Métodos para mostrar la información (diferentes gráficas y tablas)
- ✓ Complejidad para determinar la causa de los patrones específicos de aprovechamiento
- ✓ Interpretación adecuada de los datos grupales
- ✓ Interpretación adecuada entre grupos, a través del tiempo.

Todos estos aspectos, aproximaciones para analizar datos y demás recomendaciones se pueden generalizar a cualquier contexto en el cual se requiera

interpretar los resultados de la ejecución de examinandos en una prueba, con la finalidad de proporcionar información para la toma de decisiones educativas.

3. 6. Esquema informático

Tras los análisis efectuados a partir de los resultados obtenidos de la ejecución de los examinandos, resulta necesario generar para cada destinatario un informe de resultados que describa la ejecución referida al dominio del currículum y la referida a las variables contextuales. Para ello, se deben elaborar tablas, gráficas y figuras que apoyen una descripción clara, sintética y significativa de los resultados más relevantes que son de interés para cada destinatario de la información. Lo más importante de este punto es que el tipo de información que se debe describir debe ser comprensible tanto para los que emiten el informe, como para los destinatarios del mismo; y siempre debe reflejar una comparación entre los resultados y los objetivos curriculares del ámbito educativo donde se realizó la evaluación. Comúnmente existen dos formas generales de presentar los informes de resultados: mediante un informe escrito que se entrega a los usuarios; o en la forma de un documento virtual que se hace llegar al usuario mediante correo electrónico o que se publica en la página Web de la dependencia.

Capítulo IV.- PROCEDIMIENTOS PARA CAPACITAR AL CARIRE EN EL ANÁLISIS DE RESULTADOS Y LA ELABORACIÓN DE INFORMES.

En este apartado se describe la metodología que se siguió para llevar a cabo el diseño del módulo de capacitación y asesoría en torno al análisis de resultados y elaboración de informes, derivados de la evaluación criterial a gran escala. En primer término se presentará una descripción de la población para la cual se diseñó el módulo, seguida por los materiales e instrumentos que son indispensables para llevar a cabo la capacitación y la evaluación de la operación del módulo. Finalmente, se describirán los procedimientos seguidos para la elaboración del manual técnico, el diseño de formatos para el análisis de reactivos y elaboración de informes, la elaboración de ejemplos de aplicación; así como el diseño del módulo de capacitación, y la asesoría y realimentación a las actividades y productos elaborados por los miembros del Comité para el Análisis, Reporte e Interpretación de Resultados del Examen (CARIRE).

4. 1. Participantes

Los participantes serán los miembros del Comité para el Análisis, Reporte e Interpretación de Resultados del Examen (CARIRE), quienes son los responsables institucionales del análisis, reporte e interpretación de los resultados del examen obtenidos tras su aplicación a gran escala. De hecho se considera que, dado su

mejor contexto, los integrantes serán en su mayoría los miembros del Comité Coordinador del Examen, quienes acompañaron y coordinaron las actividades de los demás comités de examen durante el desarrollo del instrumento.

4. 2. Materiales e instrumentos

Para llevar a cabo la capacitación y asesoría al CARIRE se elaboraron los siguientes materiales e instrumentos:

- Un manual técnico que incluye aspectos conceptuales y metodológicos básicos para el análisis de datos derivados de la aplicación a gran escala de los exámenes criterios elaborados, así como para la elaboración de informes de resultados del examen (ver anexo 1).
- Formato general para elaborar el informe de resultados del examen (ver anexo 2).
- Ejemplos para ilustrar la aplicación de los procedimientos y criterios revisados tanto para el análisis de datos, como para la elaboración de los informes de resultados (ver anexo 3).
- Un cuestionario para evaluar la operación de los materiales producidos y de los procedimientos operados para la capacitación, asesoría y realimentación que serán proporcionadas; así como el correspondiente documento de justificación de dicho cuestionario que indica la relación entre las preguntas y

los factores evaluativos que se juzgó importante incluir en el instrumento (ver anexos 4 y 5).

- Dos listas de cotejo para realimentar los análisis e informe elaborados por los participantes (ver anexos 6 y 7).

4. 3. Procedimientos

Para la elaboración de los materiales e instrumentos antes descritos, se llevaron a cabo las siguientes acciones:

1. Elaboración de un manual técnico que incluye información acerca de los estándares para analizar los datos derivados de la aplicación de las pruebas y sobre los procedimientos necesarios para llevar a cabo el análisis de datos y la elaboración de informes especializados, según las necesidades de los destinatarios de la información; datos sobre diversos indicadores que deben conocer los destinatarios, búsqueda de información específica relevante dependiendo del uso que se le dé a los datos, métodos de análisis de información e interpretación de la misma, y referencias técnicas de diversas fuentes respecto a la elaboración de informes.
2. Diseño de formatos para registrar los resultados de la aplicación a gran escala a fin de efectuar el análisis de reactivos ya con suficientes datos empíricos, así como para consignar los resultados de los análisis de la ejecución de los examinandos en las áreas curriculares que explora el examen, y aquellos

necesarios para registrar el tipo de información requerida para la elaboración de informes, según los destinatarios de la misma.

3. Elaboración de ejemplos de aplicación de las técnicas de análisis de datos y de elaboración de informes de resultados. En ellos se ilustran diferentes procesos analíticos de casos reales en los cuales se aplicaron ciertos aspectos de la metodología, y a partir de los cuales se crearon informes de resultados. Los ejemplos presentados incluyen informes dirigidos a distintos destinatarios, a partir de los mismos resultados.

Por su parte, para proporcionar la capacitación y asesoría al CARIRE, se perfilaron los siguientes procedimientos:

1. Diseño del módulo de capacitación teórico - práctico para analizar datos y elaborar informes especializados. Se consideraron los principales temas conceptuales y metodológicos sobre el análisis de los datos derivados de las pruebas y la elaboración de informes de resultados. La capacitación se diseñó para ser impartida de manera presencial con una duración de diez horas, de conformidad con los criterios establecidos por la norma técnica del CONOCER: CRCH0542.01 "Diseño e impartición de cursos de capacitación".
2. Diseño de la asesoría y realimentación a las actividades y a los productos elaborados mediante ellas por los miembros del CARIRE. Se considera que estas acciones deben operar principalmente vía foro asincrónico y correo electrónico, para lo cual se requiere utilizar las listas de cotejo elaboradas.

3. Diseño de la evaluación de los materiales elaborados y la operación de los procedimientos utilizados, mediante la aplicación de un cuestionario a los participantes, así como mediante las listas de cotejo para los análisis de datos efectuados y los informes de resultados elaborados por los participantes.

Cabe mencionar que la planeación, diseño y elaboración de estos materiales, instrumentos y procedimientos son el producto principal de este estudio, por ello es muy importante que los participantes del módulo cuenten con los materiales e instrumentos adecuados que faciliten su capacitación, asesoría y realimentación. En este sentido, es necesario que su diseño y elaboración vigile aspectos tales como pertinencia, relevancia, grado de complejidad, rigor conceptual, estructura y claridad, entre otros.

Capítulo V.- CONCLUSIONES Y RECOMENDACIONES PARA LA OPERACIÓN Y EVALUACIÓN DEL MÓDULO DE CAPACITACIÓN Y ASESORÍA

Dentro de los objetivos primordiales de este estudio se planteó el desarrollo de procedimientos, instrumentos y materiales que permitieran cumplir con la función de capacitar y asesorar en el análisis de datos y la elaboración de informes técnicos de resultados, derivados de la aplicación a gran escala de instrumentos de evaluación del aprendizaje.

Para cumplir dicho objetivo se diseñaron los procedimientos requeridos para efectuar el análisis de datos derivados de la aplicación de pruebas criterioles de gran escala, así como los procedimientos necesarios para la elaboración de informes de resultados especializados, según las necesidades de información de diferentes destinatarios.

Se elaboró un manual técnico, el cual es el elemento crítico y esencial en el logro de los objetivos planteados, ya que éste es el producto principal de este estudio, y conforma la principal fuente de información para llevar a cabo la capacitación, y que servirá como guía metodológica y conceptual en el análisis de datos derivados de la evaluación, y en la elaboración de los correspondientes informes de resultados.

Asimismo, en el desarrollo de este estudio se crearon formatos especializados y ejemplos de aplicación para apoyar la capacitación, asesoría y realimentación de

los productos del análisis de datos, y de los procedimientos para la elaboración de informes de resultados; y finalmente se diseñaron instrumentos para evaluar la operación de los procedimientos y materiales elaborados.

Por diversas razones no fue posible operar y evaluar el módulo para capacitar, asesorar y realimentar los trabajos del CARIRE. En consecuencia, a continuación se presentan algunas sugerencias para poner en práctica los materiales, instrumentos y procedimientos que fueron desarrollados a lo largo del presente trabajo, los cuales quedarán a consideración de futuros estudios relacionados con el tema en cuestión.

- Resulta necesario seleccionar o elaborar ejemplos específicos que resulten más cercanos y significativos para los destinatarios reales de los informes de resultados del examen desarrollado. Las muestras que se presentan en el manual únicamente tuvieron el propósito didáctico de ilustrar la aplicación de los procedimientos y criterios revisados tanto para el análisis de datos, como para la elaboración de los informes de resultados.
- Se recomienda crear una aplicación de cómputo para analizar psicométricamente los ítems y modelos de examen, y otra más para analizar de manera automática la ejecución de los examinandos por área y línea de formación curricular evaluada, y elaborar los informes de resultados correspondientes. De hecho, este fue uno de los objetivos particulares que se propusieron inicialmente en este estudio, pero que no fue posible lograr.

- En este documento se elaboró un cuestionario para evaluar la funcionalidad de los materiales diseñados y la operación de los procedimientos utilizados. Sin embargo, la evaluación puede ahondar en la indagación de aspectos que es necesario mejorar, mediante una entrevista con miembros del CARIRE que reporten críticas, observaciones, limitaciones o sugerencias para mejorar los materiales y procedimientos del módulo.
- Se sugiere elaborar una estrategia para el análisis de contenido, apropiada para describir las características de los documentos que elaboren los miembros del CARIRE, mismos que contienen los análisis de datos efectuados y los informes de resultados elaborados por ellos.

El producto más importante que se obtuvo en este estudio fue un manual técnico que apoyara la capacitación y asesoría para llevar a cabo un análisis, reporte e interpretación de resultados de pruebas criterioles de gran escala. Aunque se cumplió con los objetivos planteados inicialmente, aún quedan acciones por realizar en torno al desarrollo y mejoría de la metodología propuesta. En este sentido, serán necesarios estudios de aplicación de los instrumentos, materiales y procedimientos desarrollados, dentro de escenarios reales; de tal manera que se constate su generalizabilidad a cualquier tipo de contexto educativo e institucional. En ellos, deberá tomarse en cuenta que los productos de este estudio son sensibles a las necesidades particulares de cada institución; sin embargo, por cuestiones de tiempo

y espacio, no fue posible considerar todas las posibilidades de organización curricular existentes, así como de estructura administrativa institucional. En consecuencia, será trabajo de cada institución educativa adaptar la información a las necesidades específicas que tenga.

Bibliografía /Referencias

Abad, F. J. ; Atencia, J.; García, C.; Hontangas, P.; Olea, J.; Ponsoda, V.; Revuelta, J.; Suero, M.; y Ximénez, C. (s/f). Establecimiento de los puntos de corte. Proyecto de innovación docente: Ayuda a la creación de exámenes. Departamento de psicología social y metodología. Facultad de Psicología. Universidad Autónoma de Madrid. Consultado el 14 de agosto de 2007 en: <http://www.uam.es/docencia/ace/26-corte.htm>

Aiken, L. (1996). Tests psicológicos y evaluación. Octava edición. México: Prentice Hall, pp. 3-20, 135-170

APA (1999). Standards for educational and psychological testing. Prepared by a joint committee of the American Psychological Association, American Educational Research Association, National Council on Measurement in Education. Washington, D.C.: American Psychological Association pp. 1-169.

Berk, R. A. (Ed.). (1984). *A guide to criterion – referenced test construction*. Baltimore: The Johns Hopkins University.

- Berk, R. (1986). A Consumer's Guide to Setting Performance Standards on Criterion-Referenced Tests. Review of Educational Research. Vol. 56, no. 1, pp. 137-172.
- Condelli, L. (2005). Demonstrating Results: developing state and local report cards for adult education. American institutes for research. Washington, DC. Guía preparada para el proyecto: Enhancing Performance Through Accountability. Contract # ED-04-CO-0025/0001. U.S. Department of Education, Office of Vocational and Adult Education, Division of Adult Education and Literacy.
- Consejo Asesor Externo del Centro Nacional de Evaluación para la Educación Superior (2000). Estándares de Calidad para Instrumentos de Evaluación Educativa. pp. 1-51. México: CENEVAL.
- Consejo de Normalización y Certificación de Competencia Laboral (2002). Normas técnicas de Competencia laboral: CRCH0542.01 Diseño e Impartición de Cursos de Capacitación. Recuperado en abril de 2005 en: <http://www.excela.com.mx/imagenes/Pdf/CRCH0542.01.pdf>
- Contreras L. A. (2000). Desarrollo y Pilotaje de un Examen de Español para la Educación Primaria en Baja California. Trabajo de Tesis presentado para la obtención del grado de Maestro en Ciencias Educativas en el Instituto de

Investigación y Desarrollo Educativo de la Universidad Autónoma de Baja California.

Contreras. L. A. y Backhoff, E. (2004). Metodología para elaborar exámenes criterios alineados con el currículo. En: Castañeda, S. Educación, aprendizaje y cognición. Teoría en la práctica. México: Manual Moderno. Pags. 155 - 174

Fraustro, S. M. (2000). CONOCER: Consejo de Normalización y Certificación Laboral. En: Ingenierías, Enero-marzo 2000, Vol. III, No. 7 pp. 52-57.

Gómez A., B. (1990). Evaluación criterial: una metodología útil para diagnosticar el nivel de aprendizaje de los alumnos. Madrid, España: Narcea, D. L. 55 pp.

Hambleton, R. K., y Eignor, D. R. (1980). Competency test development, validation and standard setting. En R. M. Jaeger & C. K. Tittle (eds.) Minimum achievement testing: motives, models, measures and consequences (pp. 367-396). Berkley, Ca.: Mc Cutchan.

Hambleton, R. K. (1988). Criterion-referenced measurement. Educational research, methodology, and measurement. An international Handbook. Keeves, J. P. Edit. Australia: 1988. pp. 277-281.

- Jaeger, R. M., (1989). Certification of student competence, Educational measurement. Edit. Linn, R. Estados Unidos: American council on Education and the Oxy Press. pp. 485-514.
- Jaeger, R. M. y Tucker, C. G. (1998). Analyzing, disaggregating, reporting, and interpreting student's achievement test results: a guide to practice for Title I and beyond. Estados Unidos: Council of Chief State School Officers
- Klein, T. W. (1990). Characteristics which differentiate Criterion-Referenced from Norm-referenced Tests. Nevada Department of Education at Carson City, Nevada. (ERIC No. ED 324 327).
- Livingston, S. y Zieky, M. (2004). Extractos de Puntuaciones para Aprobar. Manual para establecer normas de rendimiento referentes a exámenes académicos y ocupacionales. Estados Unidos: Educational Testing Service pp. 3-13.
- Márquez, M. C. E. (2005). Aplicación y validación de un método para determinar el punto de corte del Examen de Egreso de Inglés de la Universidad Autónoma de Baja California. Trabajo de Tesis presentado para la obtención del grado de

Maestra en Ciencias Educativas en el Instituto de Investigación y Desarrollo Educativo de la Universidad Autónoma de Baja California.

Nitko, A. J. (1984). Defining "Criterion- referenced Test". En Berk, R. A. (Ed.). A guide to criterion-referenced test construction. Baltimore: The Johns Hopkins University Press pp. 8-28.

Nitko, A. J. (1993). Designing tests that are integrated with instruction. Educational measurement. Edit. Linn, R. Estados Unidos: American council on Education and the Oxy Press. pp. 447-474.

Nitko, A.J. (1994). A Model for Developing Curriculum-Driven Criterion-Referenced and Norm-Referenced National Examinations for Certification and Selection of Students. Ponencia presentada en la Conferencia Internacional sobre Evaluación y Medición Educativas, de la Asociación para el Estudio de la Evaluación Educativa en Sudáfrica (ASSESA).

Popham, J. (1990). Modern Educational Measurement. A Practitioner's Perspective (2nd Edition). Needham Heights, MA: Allyn y Bacon p. 343.

- Popham, W. J. (2000). Modern educational Measurement: practical guidelines for educational leaders (3rd Edition). Needham Heights, MA: Allyn y Bacon pp. 64-116.
- Secretaría de Educación Pública (2001). Programa Nacional de Educación 2001-2006. Por una educación de buena calidad para todos. Un enfoque educativo para el siglo XXI. Primera edición, México: Secretaría de Educación Pública. pp.100-184.
- Universidad Autónoma de Baja California (2003). Plan de Desarrollo Institucional 2003-2006 / Universidad Autónoma de Baja California. México: UABC.
- Universidad Autónoma de Baja California (2004). Convocatoria para el Diseño de Instrumentos de Evaluación Colegiada. México: UABC
- Ward, A., Stoker, H. y Murria, M. (1996). Educational Measurement. Origins, Theories and explications. U. S. A.: University Press of America, Inc. Vol. I, pp. 1- 88, 125-134.
- Woolfolk, A. y McCune, L. (1983). Psicología de la educación para profesores. Madrid: Narcea.

ANEXOS

Anexo 1

Manual del participante del Módulo 5: Análisis, Reporte e Interpretación de Resultados de Pruebas Criteriales de Gran Escala.



Universidad Autónoma de Baja California
Instituto de Investigación y Desarrollo Educativo

Curso-Taller para la
Elaboración de Exámenes Criteriales

Manual del participante

Módulo 5.

Análisis, Reporte e
Interpretación de Resultados
de Pruebas Criteriales de
Gran Escala.

Citlalli Sánchez Álvarez

Septiembre de 2007

Objetivo general

Al finalizar el presente módulo del curso-taller, los participantes analizarán los resultados del examen y elaborarán los correspondientes informes de resultados, según las necesidades de los diferentes usuarios de la información, de conformidad con las normas técnicas.

Índice de contenidos

	Página
Objetivo general.	2
Introducción al módulo.	4
1. Introducción.	6
1.1. Análisis de resultados y elaboración de informes dentro del proceso general para desarrollar el examen	6
1.1.1. Descripción del proceso general para elaborar el examen	7
1.1.2. Descripción de los productos desarrollados por los comités de examen	8
1.2. Análisis de resultados y elaboración de informes dentro del proceso de evaluación del aprendizaje	11
1.2.1. Propósitos de la evaluación del aprendizaje y de los informes de resultados	11
1.2.2. Esquema Analítico- Informático para el procesamiento de los resultados de la aplicación a gran escala de un examen criterial	13
2. Procedimientos para analizar los resultados del examen	16
2.1. Descripción del proceso general para realizar un análisis de reactivos	16
2.2. Análisis de la ejecución de los examinandos en el examen	18
2.2.1. Análisis de la ejecución de los examinandos por área del currículum.	27
2.2.2. Análisis de la ejecución de los examinandos asociada con variables de contexto	29
2.3. Establecimiento de estándares de ejecución y puntos de corte	33
3. Procedimiento para elaborar informes de resultados.	41
3.1. Desarrollo de informes de resultados	41
3.2. Tipos de destinatarios de la información	44
3.3. Análisis, reporte e interpretación de resultados en forma desagregada	51
3.4. Ejemplos de informes de resultados de exámenes criterios de gran escala	61
Conclusiones generales	71
Glosario de términos sobre la evaluación del aprendizaje	73
Bibliografía	80

Introducción al módulo

Al finalizar el presente módulo del curso-taller para la elaboración de exámenes criterios, se espera que el participante haya efectuado un análisis de los resultados obtenidos tras la aplicación a gran escala de los modelos de examen previamente elaborados y que haya elaborado los correspondientes informes de resultados, según los destinatarios de la información. Para lograrlo, el enfoque del módulo será el de un curso-taller en el cual se revisarán primero aspectos conceptuales y procedimentales relativos a los criterios técnicos que será necesario atender para analizar los datos y elaborar los informes. Posteriormente, lo aprendido se aplicará en la modalidad del taller para lo cual se contará con la asesoría y realimentación que proporcionará un instructor especialista en asuntos psicométricos.

Los temas que se revisarán y la secuencia que se seguirá al tratarlos serán los siguientes: primero será necesario ubicar el análisis de resultados y la elaboración de informes dentro del proceso general seguido para construir el examen; es decir, se describirán las etapas y procedimientos empleados por los comités diseñador, elaborador de especificaciones de reactivos y elaborador de ítems para obtener sus correspondientes productos. Posteriormente se describirán los propósitos de la evaluación del aprendizaje a gran escala y la función que tienen en ella el análisis de la ejecución de los examinandos y los informes de resultados; se describirá además la estructura general de los informes, así como los tipos de

destinatarios y niveles de información que deben incluirse para satisfacer sus necesidades y apoyar sus decisiones. También se revisarán los aspectos conceptuales y procedimientos psicométricos necesarios, tanto para el análisis de los ítems y modelos de examen, como para el análisis e interpretación de los resultados de la ejecución de los examinandos en la prueba. Finalmente se revisarán y aplicarán los procedimientos necesarios para elaborar diferentes tipos de informes, enfocados a satisfacer distintas necesidades institucionales.

Debido a lo anterior, es requisito que el participante de este curso haya transitado por los cuatro procesos anteriores de diseño y elaboración del examen, de tal manera que la información contenida en este curso le resulte natural y esté contextualizado.

1. Introducción.

Objetivo particular: los participantes ubicarán su intervención en el proceso de desarrollo del examen.

En esta unidad se espera que el participante ubique su intervención dentro del proceso general para construir la prueba y que revise con algún detalle los materiales que fueron elaborados por los comités que antecedieron su trabajo; especialmente los productos desarrollados por el Comité Diseñador del Examen. Posteriormente, se revisarán los propósitos, la estructura y el nivel de especificidad apropiado tanto del análisis de los datos como de la elaboración de los informes de resultados. Así, el enfoque de los temas será básicamente conceptual y contextual, con lo cual el participante logrará una comprensión cabal de las tareas que le corresponde realizar.

1.1. Análisis de resultados y elaboración de informes dentro del proceso general para desarrollar el examen

Objetivos específicos:

- Los participantes ubicarán los procedimientos para el análisis de resultados y la elaboración de informes, como parte de las etapas correspondientes al análisis primario y secundario de resultados de la prueba.
- Los participantes conocerán los productos elaborados por los comités que construyeron los exámenes, mismos que serán la base del trabajo que les corresponde realizar.

1.1.1. Descripción del proceso general para elaborar el examen

El proceso general que se siguió para construir el examen considera las etapas de trabajo y procedimientos que se describen en la tabla 1, misma que se ilustra a continuación:

Tabla 1. Proceso general para construir el examen

Etapas	Procedimientos
1. Definir el dominio de resultados que pretende el currículum	- Selección y capacitación del Comité Diseñador del examen
	- Análisis del contenido curricular del área
	- Análisis complementario
2. Analizar el currículum	- Elaboración de la retícula del contenido a evaluar
3. Desarrollar un plan de evaluación	- Muestreo de resultados de aprendizaje a evaluar
	- Diseño de especificaciones de ítems
	- Capacitación del Comité Elaborador de ítems
4. Producir y validar ítems	- Elaboración de ítems según las especificaciones
	- Revisión de la congruencia ítem-especificación
	- Ensayo empírico y revisión de ítems
	- Revisión de la prueba y estructuración de una muestra de ítems representativa del dominio curricular
	- Ensayo empírico de gran escala y revisión de ítems
5. Análisis primario	- Especificación de estándares de calidad para el contenido del ítem, para la calidad técnica del ítem y para la calidad integral de los puntajes del examen
	- Diseño y aplicación de procedimientos para calificar el examen y analizar los resultados
6. Análisis secundario	- Análisis de resultados para elaborar los reportes de resultados por escuela y municipio, por eje temático del plan de estudios y por destinatario

Como puede observarse en la tabla, el trabajo a desarrollar en este módulo por los miembros del Comité para el Análisis, Reporte e Interpretación de Resultados del Examen (CARIRE) se refiere a los procedimientos correspondientes a la quinta etapa y al procedimiento de análisis de resultados para la elaboración de informes de

la sexta etapa. Ello significa que hasta el momento se han realizado las actividades correspondientes a las cuatro primeras etapas y que con ello se han obtenido ya los productos necesarios para que los miembros del CARIRE puedan realizar las acciones que les serán encomendadas en el presente módulo, mismas que por su complejidad constituyen sin duda un reto intelectual que pondrá a prueba su capacidad profesional y docente, así como su creatividad.

1.1.2. Descripción de los productos desarrollados por los comités de examen

Los productos que fueron elaborados por los miembros de los comités que desarrollaron el examen en los cuatro módulos anteriores del curso-taller, se describen a continuación:

Retícula del curso. La retícula es un modelo gráfico que muestra los contenidos del área, materia o curso; o en su caso, del eje, subeje y línea de formación definidas por el currículum, y las relaciones de servicio que dan y reciben dichos contenidos. En la retícula, los contenidos están identificados mediante rectángulos y las relaciones entre ellos mediante flechas. El número de flechas que surgen de cada contenido o que llegan a él representa su función principal dentro del programa de estudios; así, tenemos que la función de un contenido fuente es dar servicios a otros contenidos; la de un contenido sintético es recibir e integrar servicios de otros contenidos; un contenido aislado no da ni recibe servicios de otros; y una rama es aquel contenido que se caracteriza por dar y recibir servicios de contenido. Si el rectángulo está enmarcado por otro rectángulo o tiene un color especial o algún otro rasgo distintivo

según lo indique el código de la retícula, significa que tal contenido se juzgó como esencial o de primera importancia en el contexto del programa.

Documento de justificación del contenido importante a evaluar. Se trata de un escrito que presenta las decisiones adoptadas respecto a la importancia relativa de los contenidos, los argumentos por los cuales fueron juzgados como esenciales o como importantes en el contexto del currículum y las resoluciones adoptadas respecto a ellos: censarlos en el examen por ser esenciales, o incorporarlos a la prueba por ser importantes y por haber sido seleccionados al azar de entre todos los importantes. Al respecto cabe señalar que tales decisiones son adoptadas en virtud de que todo examen criterial es una muestra pretendidamente representativa del dominio curricular, dada la imposibilidad de evaluar en una sola ocasión todo el contenido juzgado como importante.

Tabla de especificaciones del examen. Se trata de una tabla de doble entrada que resume las decisiones adoptadas en los dos documentos antes descritos, respecto a la planeación de la prueba. Especifica el contenido cuyo dominio será evaluado, el número de especificaciones de reactivos que se requieren para evaluarlo, así como el número y tipo de reactivos que será necesario construir según cada especificación. De hecho, la tabla de especificaciones presenta la estructura general del examen. Cabe señalar que el número de especificaciones y de reactivos que se indican en la tabla, fue ponderado con base en criterios técnicos como el tipo de estimación de la ejecución del examinando y el tipo de impacto de la prueba (Popham, 1990), así como a partir de criterios operativos como el acuerdo de considerar en dos horas la

duración de una sesión de examen y considerar que un reactivo de opción múltiple se responde en aproximadamente entre un minuto y un minuto y medio.

Documento del análisis psicométrico de reactivos y modelos piloteados. Se trata del registro obtenido tras la prueba empírica de los ítems y modelos de examen ante una muestra de examinandos de pequeña escala, mismo que contiene una relación de indicadores psicométricos obtenidos después del pilotaje, entre los cuales se encuentran la dificultad de cada reactivo, su discriminación y las respuestas de los examinandos ante cada opción, así como el índice de confiabilidad de cada modelo de examen.

Ítems y modelos de examen depurados tras el pilotaje. Se trata de las versiones finales del examen que fueron aplicadas a la muestra de gran escala. Para obtenerlas fue necesario revisar los ítems y modelos a partir de los resultados de los análisis realizados mediante el jueceo y mediante el ítem análisis¹; es decir, se modificaron o sustituyeron los reactivos que presentaban sesgo, demanda cognitiva excesiva, limitado alineamiento con el currículum, dificultad inadecuada, discriminación por debajo del criterio establecido o los que presentaban opciones de respuesta que no fueron seleccionadas por algún examinando, entre otros problemas técnicos.

En conclusión, podemos decir que los procedimientos de análisis de la información derivada de la aplicación a gran escala de los modelos de examen

¹ Para objeto de este documento, a partir de este momento se utilizará el término “análisis de reactivos” para referirse al ítem análisis.

desarrollados y la elaboración de los correspondientes informes, constituyen las fases finales de un proceso que hasta ese momento abarca desde la planeación de la prueba y su desarrollo, hasta su prueba empírica a gran escala. Para que el CARIRE pueda cumplir con tal función dual, se requiere efectuar una síntesis tanto de los resultados del jueceo como los del análisis de reactivos realizado tras el pilotaje, así como de los tres elementos de la planeación de la prueba que fueron aportados por el comité diseñador del examen: la retícula donde se estructura el contenido del curso, el documento de justificación del contenido importante a evaluar y la tabla de especificaciones del examen.

1.2. Análisis de resultados y elaboración de informes dentro del proceso de evaluación del aprendizaje

Objetivo. Los participantes ubicarán el análisis de resultados y la elaboración de informes dentro del proceso de evaluación del aprendizaje.

1.2.1. Propósitos de la evaluación del aprendizaje y de los informes de resultados

En el ámbito educativo se tiene bien establecido que la función primordial de la evaluación del aprendizaje que logran los alumnos es reportar los resultados a todos aquéllos individuos o instancias involucradas, con la finalidad de que se vean mejorados los procesos educativos. Cuando un informe de resultados se encuentra debidamente elaborado, representa un apoyo fundamental para todas aquéllas

iniciativas que buscan el mejoramiento de los programas educativos; pero si ocurre lo contrario, puede generar interpretaciones incorrectas de la información, que a su vez afectan en forma negativa a los alumnos, maestros y a la comunidad escolar en general (Roeber, 1995).

Glasser (citado por Hambleton, 1988), publicó en 1963 un artículo que menciona algunos métodos de evaluación que proporcionan la información necesaria para la toma de decisiones en relación a objetivos y competencias específicas. Por su parte, Popham (1978, citado por Hambleton en 1988), señala que el objetivo de las pruebas referidas a un criterio es permitir la interpretación de la ejecución de los examinandos en relación a competencias que previamente fueron establecidas. Esto es de suma importancia, ya que así es más fácil elaborar ítems que correspondan en forma directa a dichas competencias, así como llevar a cabo una interpretación adecuada de los resultados de la ejecución de los examinandos.

Por ello, debe establecerse un estándar para cada criterio o área que mida la prueba, ya que se evaluará la ejecución de los examinandos en relación a dichos criterios y posteriormente se llevará a cabo una interpretación congruente con los resultados de la prueba.

En este sentido, la información que proveen las pruebas criterioales es de gran valor para los destinatarios. Cabe mencionar que los resultados de estas pruebas tienen tres principales usos (Hambleton, 1988):

- 1) Describir la ejecución de los examinandos

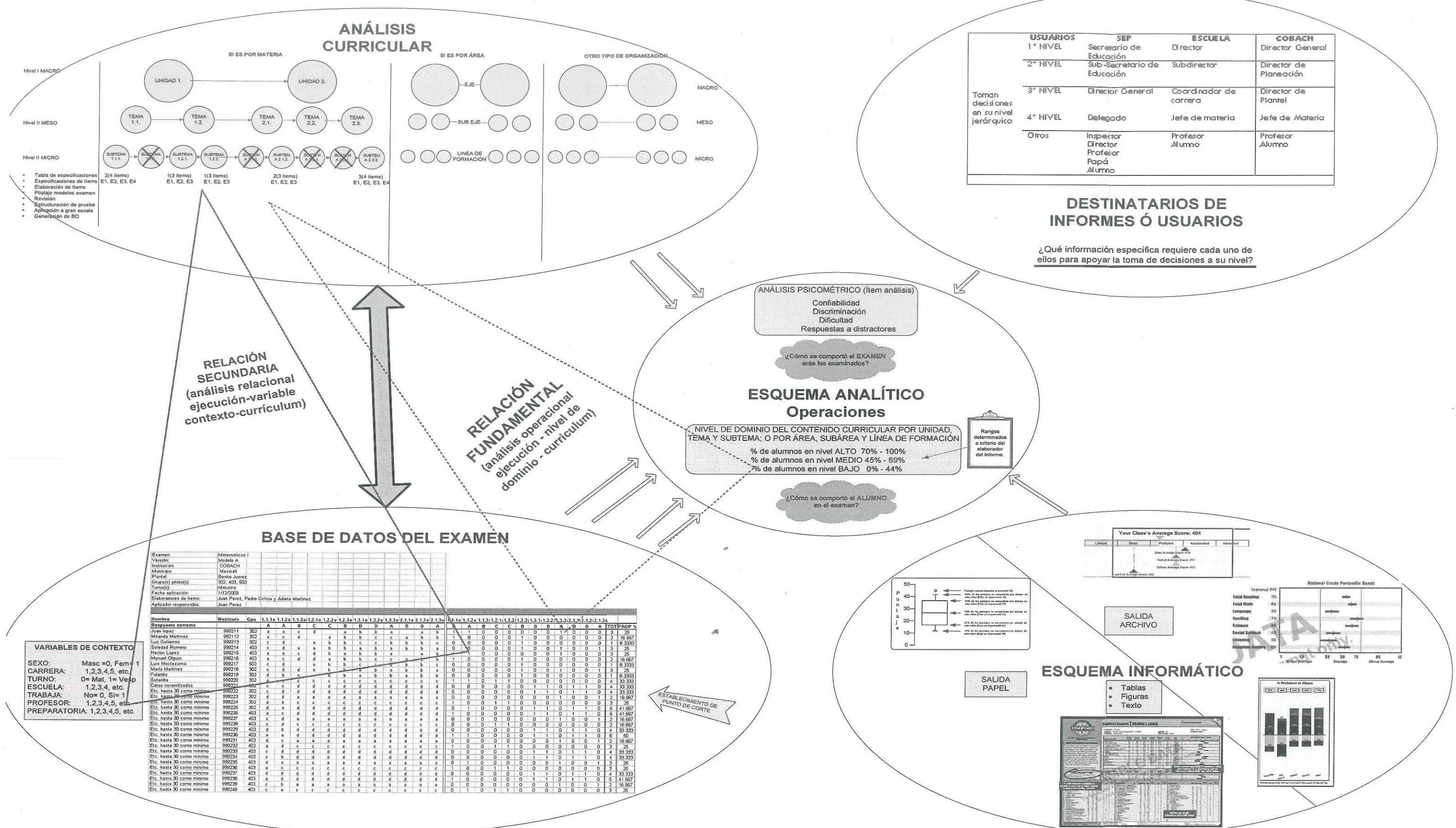
- 2) Determinar el nivel de dominio de los examinandos (Por ejemplo, diferenciar entre aquellos alumnos que tienen el nivel de competencia y los que no lo poseen), y
- 3) Describir la ejecución de grupos específicos de examinandos, en los estudios de evaluación de programas.

Así, los informes de resultados tienen como propósito principal proporcionar a los usuarios los elementos necesarios para comprender el uso que se le dará a la información como parte de un proceso de mejoramiento escolar. Para cumplir con dicho propósito, el CARIRE deberá realizar las siguientes acciones generales:

- Determinar desde un inicio las metas específicas que persigue el informe de resultados.
- Seleccionar las estrategias que sean consistentes con dichas metas, y que proporcionen de manera efectiva la información requerida.
- Comunicar los resultados de la evaluación en forma clara, precisa y significativa a los destinatarios de la información.

1.2.2. Esquema Analítico- Informático para el procesamiento de los resultados de la aplicación a gran escala de un examen criterial.

En la figura 1 se presenta un esquema que contiene los principales elementos y relaciones a considerar para desarrollar, tanto los análisis de los datos obtenidos tras aplicar el examen, como los correspondientes informes de resultados.



De manera general, este esquema se encuentra conformado por los siguientes elementos, procesos y relaciones que se dan entre ellos:

- **Análisis curricular.** Representado por la retícula que se construyó para identificar y mostrar la estructura del contenido que evalúa el examen. Así, los ítems de la prueba exploraron el dominio de los contenidos que se juzgó importante evaluar.
- **Base de datos del examen.** Representa la ejecución de los examinandos en la prueba; es decir, las respuestas que dieron a los ítems.
- **Variables de contexto.** Presenta condiciones propias de los examinandos, el hogar, la escuela o de la operación del currículum que pudieron estar relacionadas con su ejecución en la prueba.
- **Esquema analítico.** Se refiere a operaciones como la estimación de la calidad del examen mediante el análisis de reactivos o la determinación del nivel de dominio que tuvieron los examinandos sobre el currículum.
- **Destinatarios de los informes.** Considera los usuarios de la información del examen y el tipo de información más apropiada para ellos, según el nivel de decisiones que les corresponde adoptar.
- **Esquema informático.** Considera los tipos de salida de la información del examen para los usuarios y los tipos de datos textuales o gráficos que se les presentarán, a fin de que resulten comprensibles y significativos para sus necesidades.

Más adelante, en el tema 2.2, se describirán con mayor detalle los elementos y relaciones que integran el esquema analítico – informático; en particular, los procesos que requiere el análisis de los datos de la prueba.

UNIDAD 2. Procedimientos para analizar los resultados del examen.

Objetivo particular: los participantes conocerán los aspectos conceptuales y metodológicos básicos involucrados en el análisis de resultados del examen.

Introducción a la Unidad

En esta unidad se revisarán los propósitos del análisis de la información derivada de la aplicación del examen a gran escala, así como los aspectos conceptuales y los procedimientos específicos que se utilizan para llevar a cabo tanto el análisis psicométrico de ítems y modelos, como para analizar la ejecución de los examinandos en la prueba y establecer estándares y puntos de corte; todo ello a fin de estar en condiciones de elaborar los informes de resultados correspondientes.

2.1. Descripción del proceso general para realizar un análisis de reactivos

El proceso de generación de informes de resultados del examen debe ser precedido por un riguroso análisis de resultados psicométricos, también llamado análisis de reactivos, mismo que se efectúa en esta etapa de la construcción de la prueba, cuando se cuenta ya con un número suficiente de datos empíricos derivados de la aplicación a gran escala de los modelos de examen. La idea es obtener evidencias sobre la calidad técnica de la prueba a la luz de los estándares de calidad definidos para el examen. Para ello, se seguirán los mismos procedimientos que fueron descritos en el tercer módulo, cuando se llevó a cabo un análisis de reactivos

para efectuar la calibración de los ítems y modelos de examen tras su aplicación piloto a una pequeña muestra de examinandos.

Así, el análisis de los reactivos y modelos de examen se llevará a cabo utilizando los procedimientos convencionales de la teoría clásica de los tests; es decir, la determinación del índice de dificultad de los reactivos, de su índice de discriminación, y del correspondiente coeficiente de discriminación (r_{bis}), así como el cálculo del índice de confiabilidad de las versiones de examen elaboradas. También se llevará a cabo el análisis de las respuestas a los distractores, para detectar aquellos que no fueron seleccionados por al menos un mínimo de los examinandos menos capaces.

En este momento, estos análisis tienen como propósito valorar y asegurar la calidad de los ítems y modelos de examen, a la luz de estándares de calidad agrupados en dos áreas de control: calidad técnica de cada ítem y calidad de las calificaciones de la prueba. De esta manera las áreas de control de calidad, los estándares y medidas asociadas a ellos y el criterio que debe satisfacerse en cada caso son los siguientes (Nitko, 1994, adaptado por Contreras, 2000):

Tabla 2. Áreas de control, estándares, medidas y criterios, para el análisis de reactivos y modelos de la prueba.

Área de control de calidad	Estándar	Medida	Criterio que debe satisfacer
Calidad técnica de cada ítem del test	1. Dificultad apropiada 2. Discriminación apropiada	1. Valor "p" del ítem, obtenido de la muestra de ensayo 2. Índice de discriminación 3. Coef. de discriminación	1. $.05 < p < .95$ 2. $P_a - P_b > .2$ 3. $r_{bis} > .2$
Calidad de las calificaciones del test	1. Alta confiabilidad	1. Coeficiente alfa o Kuder-Richardson 20	1. Coeficiente mayor o igual que .80 en cada modelo

2.2. Análisis de la ejecución de los examinandos en el examen.

Para elaborar informes de resultados del examen, es necesario primero efectuar un análisis de la ejecución de los examinandos, por área curricular definida en la planeación de la prueba, según sean en cada caso las necesidades y tipos de decisiones educativas que adoptarán los destinatarios de la información.

En el inciso 1.2.2 se presentó una descripción general del esquema analítico – informático que aparece en la figura 1. A continuación se describen con algún detalle los principales elementos y procesos contenidos en él.

ANÁLISIS CURRICULAR.

Se trata de un análisis de contenido de fuentes formales e informales que definen el currículum, tales como los planes y programas de estudio, materiales instruccionales, prácticas educativas y otros aspectos que se consideran como representativos del universo de conocimiento que se evalúa. Este análisis se puede llevar a cabo de acuerdo al tipo de organización curricular de que se trate; por ejemplo, por unidad-tema-subtema, por eje – subeje – línea de formación, o por cualquier otro tipo de organización que presente el contenido del currículum.

Este análisis es de suma importancia, ya que la calidad de los reactivos de un instrumento de evaluación de tipo criterial es constantemente juzgada en relación a los resultados que pretende el currículum. El proceso de análisis curricular culmina con un producto notable llamado retícula, que en forma sintética y gráfica permite identificar y mostrar la estructura del contenido que evalúa el examen por niveles denominados macro, meso y micro. En la parte superior izquierda de la figura 1, se

ilustra un ejemplo bosquejado de una retícula. Otros productos que se derivan del análisis curricular, son: la tabla de especificaciones del examen, las especificaciones de cada uno de los reactivos que conforman la prueba, los reactivos elaborados a partir de dichas especificaciones y la estructuración de los ítems en los diversos modelos de la prueba que servirán para evaluar el dominio que tienen los examinandos del contenido curricular. De esta manera se cuenta con un análisis detallado del contenido importante a evaluar, así como los documentos que justifican dicho análisis.

BASE DE DATOS DEL EXAMEN

Constituida por una tabla de doble entrada que contiene varios elementos, entre los cuales se encuentran los siguientes:

- **EXAMINANDOS.** Se especifican los nombres o forma de identificar a los examinandos en los renglones de la tabla.
- **ÍTEMS.** Se presentan en las columnas de la tabla.
- **DATOS CRUDOS.** En consecuencia, cada celda contendrá la letra que corresponde a la opción (a, b, c, d, etc.; según sean tres, cuatro o más opciones de respuesta) que seleccionó cada alumno en cada reactivo de opción múltiple del examen. Éste se considera el dato primario o crudo de la prueba.
- **CALIFICACIÓN DE LA PRUEBA Y OBTENCIÓN DEL PUNTAJE TOTAL.** Posteriormente, estos datos se convertirán en datos interpretables mediante un procedimiento que inicia con la calificación de la prueba, la que a su vez

comienza con la conversión de las letras en las celdas (a, b, c, d) en ceros y unos, mediante la comparación de las respuestas de cada examinando con la clave de respuestas correctas; siendo para cada examinando, en cada reactivo: “cero” incorrecto y “uno” correcto. Enseguida, al final de la base de datos se presenta una columna con puntajes totales logrados por los examinandos, mismos que se obtienen mediante la sumatoria de los aciertos conseguidos en los ítems por cada alumno. Para ser interpretables, estos datos se deberán convertir en algún tipo de puntuación referida a un criterio. En las pruebas criterioles, el dato “duro” que se utiliza comúnmente para informar sobre los resultados de la prueba es el porcentaje de aciertos, debido a que brinda una síntesis de la ejecución en el examen que resulta familiar y permite ubicar rápidamente el nivel de dominio logrado en la prueba. Usualmente, también se presentan los porcentajes de aciertos en rangos que forman una escala ordinal, a fin de representar de manera cualitativa estándares de ejecución o niveles de dominio (por ejemplo, alto, medio y bajo) de los contenidos que miden los ítems de la prueba.

VARIABLES DE CONTEXTO.

De manera relacionada, el porcentaje de aciertos en cada área del currículum puede ser vinculado con variables contextuales del propio examinando, de la escuela, de su entorno u hogar, de las condiciones de operación del currículum y otras más, a fin de que la actuación del examinando resulte más significativa. Por ejemplo, se puede describir si el promedio de calificaciones en el ciclo anterior está

relacionado con el puntaje que obtuvo el examinando en determinada área curricular. En el inciso 2.2.2 se presentarán con mayor detalle algunas variables contextuales que usualmente se relacionan con las puntuaciones obtenidas por los examinandos. Para efectos de análisis estadístico de los datos, comúnmente se lleva a cabo una codificación de dichas variables, mediante la conversión de los datos a una escala numérica, según sea el caso. Por ejemplo, en la variable de género, se puede codificar femenino con un “cero” y masculino con “uno”; la variable de carrera con uno, dos, tres, cuatro, etc., según la diversidad de carreras que cursen los examinandos.

Así, este análisis relacional pretende descubrir las posibles relaciones entre las variables de contexto, con la ejecución del examinando en el examen, referida al dominio del currículum. En la figura 1 dichos vínculos se representan mediante un triángulo de relaciones entre los resultados en la base de datos, el análisis curricular y las variables de contexto.

ESQUEMA ANALÍTICO.

En un examen criterial alineado con el currículum queremos dar cuenta de en qué medida los resultados obtenidos por los examinandos reflejan el dominio de dicho currículum. En consecuencia, después de obtener el porcentaje total de aciertos, es necesario determinar el porcentaje de aciertos que tuvo cada examinando en los grupos de ítems que evaluaron el dominio de cada unidad y tema (o de cada categoría en que esté organizado el contenido, según la estructura del currículum). Lo anterior con la finalidad de saber qué segmentos del currículum

domina el alumno, y qué segmentos no domina. Esta relación es de primera importancia, ya que muestra el vínculo estrecho que guardan el análisis curricular y la base de datos.

En la figura 1 esta relación se representa mediante un triángulo de relaciones fundamentales que se denomina análisis operacional, pues considera las operaciones necesarias para evaluar el nivel de ejecución de los examinandos en la prueba, referido al dominio del currículum, desagregado por área que evalúa la prueba. Este análisis considera también las evidencias recabadas sobre la calidad técnica del instrumento (ver inciso 2.1), mismas que pondrán un límite a la confianza que podemos tener en los resultados de la prueba y a la validez de las inferencias que podremos llevar a cabo a partir de ellos.

DESTINATARIOS DE LOS INFORMES.

Los destinatarios de los informes de resultados son un elemento clave que sirve para determinar el nivel de profundidad y detalle con el que se elaborarán tanto los análisis de datos como el informe, así como los tipos de información que deberán incluir. Es necesario hacer una descripción detallada de las características de cada uno de los destinatarios a los cuales se dirige el informe, del tipo de información que necesitan y de los usos que le darán a la información.

Lo anterior significa que una vez que se tiene el análisis de la ejecución de los examinandos, se deben elaborar informes de resultados para los usuarios, quienes toman decisiones en su nivel jerárquico, y por ende tienen necesidades informativas diferenciales para fundamentar la toma de dichas decisiones. En consecuencia,

diversos modos de organización institucional nos proveen diferentes modelos de destinatarios.

Por ejemplo, para un funcionario de primer nivel en una institución educativa se deberá elaborar un informe sobre la ejecución de la población general que fue examinada, clasificada por municipio, carrera y unidades temáticas. Se podrá relacionar la ejecución general de los examinandos con algunas variables de contexto, tales como tipo de escuela, sexo, turno, etc. Este informe no debe ser muy detallado, pues el usuario toma decisiones generales respecto a la planeación de la institución.

A funcionarios de niveles jerárquicos inferiores (coordinadores de carrera, jefes de materia, docentes, etc.), se les proporcionará información más específica; por ejemplo, porcentajes de aciertos de los alumnos de la carrera o materia; por área curricular evaluada y de los alumnos de todos los maestros que imparten la materia, con la finalidad de que se puedan comparar y mejorar estrategias docentes.

Lo que debe destacarse es que, en los informes de resultados, independientemente del destinatario y el nivel jerárquico al que corresponda, siempre deberá incluirse la información del triángulo de relaciones fundamentales antes mencionado.

Como producto de la literatura revisada, y con el propósito de simplificar la clasificación de los destinatarios, se proponen cuatro niveles:

Primer nivel: si la evaluación es de tipo Estatal, el informe deberá dirigirse al Secretario de Educación del Estado, o al Director General de la institución. En estos casos, tanto los análisis como el informe deben incluir información respecto a:

- Distribución de promedios de las escuelas del Estado, en cada área del currículum evaluada.
- Distribución de promedios de las escuelas, por municipio, en cada área curricular.
- Tendencia del municipio y promedios estatales en cada área curricular, durante los cinco años anteriores.
- Gráficas que comparen simultáneamente los resultados de logro contra acciones educativas determinadas (perfil de profesores, recursos de la escuela, etc.)

Si la evaluación es de tipo departamental o colegiada, el informe deberá dirigirse al Director del plantel, en cuyo caso deberá incluir el mismo tipo de información, pero haciendo referencias al nivel de la escuela de que se trate.

Segundo nivel: si la evaluación es de tipo Estatal, departamental o colegiada, el informe deberá dirigirse al Sub-Secretario de Educación del Estado, al Director general de planeación educativa ó al Subdirector del plantel; en cuyo caso deberá incluir información respecto a:

- Promedio de ejecución de estudiantes en grupos de preguntas que evalúan blancos de aprendizaje claves, por área del currículum.
- Promedio de ejecución de estudiantes en cada área del currículum.

- Resultados de ambos aspectos, por municipio o, en su caso por carrera.

Tercer nivel: si la evaluación es de tipo Estatal, departamental o colegiada, el informe deberá dirigirse al Director General, Delegado educativo, Inspector escolar, Coordinador de carrera, ó Director de plantel, y deberá incluir información acerca de:

- Promedio de resultados por escuela dentro de la región de servicio, para cada área del currículum
- Promedio de resultados por grupo, en cada unidad y tema
- Tendencia de la ejecución promedio, durante los cinco años anteriores

Cuarto nivel: en este nivel se incluyen al Jefe de materia, profesores, y otros, según el tipo de evaluación. Este informe deberá incluir datos respecto a:

- Promedio general y por género, de resultados de los estudiantes de la escuela, por área del currículum evaluada.
- Porcentaje de aciertos por área del currículum, para cada examinando.
- Comparación del promedio general y por género, de los resultados de los estudiantes de la escuela, por área del currículum, contra las correspondientes medias de las escuelas del Estado.

Se considera que pueden existir otros niveles, en los cuales se encuentran destinatarios tales como los padres de familia, alumnos y público en general. Para este tipo de informes el nivel de especificidad y detalle de la información deberá ser adaptada para las características específicas y necesidades informativas que tengan.

En la tabla 4, que se muestra más adelante en el inciso 3.2, se presenta una síntesis de los destinatarios de la información de un examen criterial de gran escala y

de los tipos de información apropiados para ellos, según el tipo de decisiones que corresponden a su nivel jerárquico. En dicho inciso se revisará con mayor detalle este aspecto.

ESQUEMA INFORMÁTICO.

Tras los análisis efectuados a partir de los resultados obtenidos de la ejecución de los examinandos, resulta necesario generar para cada destinatario un informe de resultados que describa la ejecución referida al dominio del currículum y la referida a las variables contextuales. Para ello, se deben elaborar tablas, gráficas y figuras que apoyen una descripción clara, sintética y significativa de los resultados más relevantes que son de interés para cada destinatario de la información. Lo más importante de este punto es que el tipo de información que se debe describir debe ser comprensible tanto para los que emiten el informe, como para los destinatarios del mismo; y siempre debe reflejar una comparación entre los resultados y los objetivos curriculares del ámbito educativo donde se realizó la evaluación. Comúnmente existen dos formas generales de presentar los informes de resultados: mediante un informe escrito que se entrega a los usuarios; o en la forma de un documento virtual que se hace llegar al usuario mediante correo electrónico o que se publica en la página Web de la dependencia.

2.2.1. Análisis de la ejecución de los examinandos, por área del currículum

Como ya se señaló en el inciso 1.2.2, una vez aplicado el examen deberá ser calificado y posteriormente sus resultados serán analizados. Para llevar a cabo ambos procesos se requiere seguir el siguiente procedimiento:

1. Calificación de las respuestas que dieron los examinandos en la prueba. En el contexto de la evaluación a gran escala existen las siguientes dos modalidades generales para calificar la ejecución de los examinandos en el examen:

- La primera de ellas comúnmente implica leer y dar mantenimiento a las hojas de respuesta en un lector óptico y, a partir de las respuestas de los examinandos ya digitalizadas, estructurar una base de datos.
- La segunda modalidad comúnmente implica haber respondido el examen aplicado mediante una computadora, con lo cual los datos aportados por los examinandos al responder se estructuran directamente en una base de datos diseñada como parte de la aplicación que administra la prueba.

En ambas modalidades la base de datos obtenida contiene inicialmente las respuestas crudas que dieron los examinandos a cada uno de los ítems de la prueba. En el caso de los reactivos de opción múltiple, las respuestas se refieren a la letra correspondiente a la opción que fue seleccionada por cada examinando en cada reactivo.

Posteriormente, las respuestas crudas que dieron los examinandos a los ítems de cada modelo son comparadas contra la clave de respuestas correctas

correspondiente. Así, la tabla de doble entrada que contiene las letras que corresponden a las opciones elegidas por cada examinando, se convierte en una tabla que contiene en su lugar ceros y unos, según sean incorrectas o correctas sus respuestas en cada ítem.

Finalmente, se obtiene el total de respuestas correctas que logró cada estudiante mediante la suma de unos que logró en los ítems del modelo que respondió (filas de la tabla). De igual modo, se obtiene el total de respuestas correctas que lograron todos los examinandos en cada ítem del modelo (columnas de la tabla). Dicho procedimiento se ilustra en la figura 2

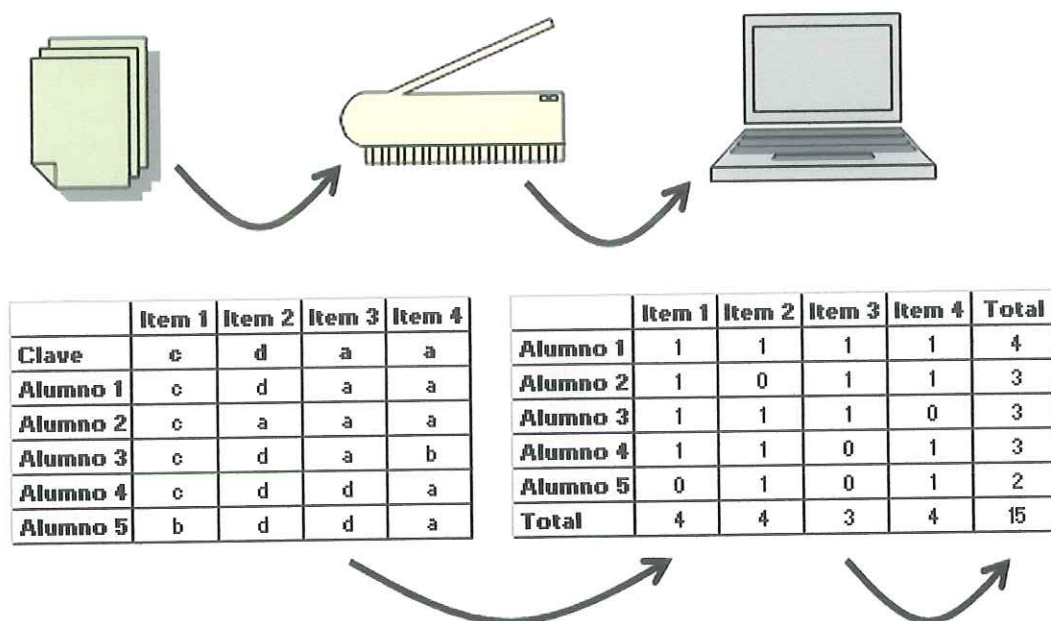


Figura 2. Ilustración del proceso para calificar la prueba

El puntaje de aciertos en cada ítem es el dato fundamental para efectuar el análisis de reactivos al que se refiere el inciso 2.1 ya comentado. Por su parte, el

puntaje total logrado por cada examinando constituye el insumo esencial para determinar el porcentaje de aciertos global que cada estudiante tuvo en la prueba, así como el porcentaje que logró en cada área del currículum que el examen evalúa. De esta forma se obtendrá información respecto a cuáles segmentos específicos del contenido domina y cuáles no domina.

2. 2. 2. Análisis de la ejecución de los examinandos asociada con variables de contexto

Para llevar a cabo este tipo de análisis es necesario asociar variables de contexto con los datos que se obtienen de la ejecución de los examinandos, de tal manera que los resultados se tornen más significativos. En este sentido, es importante considerar la relación existente entre las respuestas de los examinandos y variables tales como tipo de escuela, nivel socioeconómico del alumno, nivel escolar de los padres, género, turno, maestro que imparte la materia, si trabaja o no, etc.

En el tema 2.2 también se hizo mención de cómo el porcentaje de aciertos en cada área del currículum puede ser vinculado con variables contextuales del propio examinando, de la escuela, del hogar y de las condiciones de operación del currículum.

A continuación, se presentan algunas variables contextuales que se clasificaron como producto de la literatura revisada, y que usualmente se relacionan con las puntuaciones obtenidas por los examinandos.

- **Variables propias del alumno:** entre este tipo de variables se encuentran el género, la edad, la escuela de procedencia, si trabaja o no trabaja, el estado civil, etc.
- **Variables propias de la Escuela:** incluyen el turno, la carrera que estudia el examinando, la zona escolar a la que pertenece la escuela, el nivel educativo que se imparte en el plantel, el tipo de control o sistema al que pertenece la escuela (pública, privada, abierta, etc.), y otros.
- **Variables propias del hogar en el que vive el examinando:** este tipo de variables incluyen aspectos del entorno del examinando, tales como la escolaridad de la madre, ocupación del padre, ingreso familiar mensual, cantidad de miembros en la familia, la zona en la que vive, el tipo de vivienda, etc.
- **Variables propias de la operación del currículum:** aquí se pueden incluir el estilo didáctico de quien imparte la materia, si cuenta o no con tutor académico, la disponibilidad de libros de texto y el perfil docente, entre otros.

Cualquier variable perteneciente a alguno de los tipos mencionados en los párrafos anteriores puede ser asociada con la puntuación obtenida en la prueba. Sin embargo, como se mencionó en el inciso 2.2, debe efectuarse una codificación de cada variable pues el análisis estadístico correlacional requiere que ambas variables que se relacionan se encuentren en una escala. Cuando ello sucede, es posible establecer una medida de correlación entre la variable de contexto de interés y la

puntuación obtenida por los examinandos en la prueba, así como también el nivel de significancia estadística de dicha relación.

Para realizar este análisis correlacional puede utilizarse una variedad de programas de análisis estadístico, tales como SPSS, EZAnalyze (de distribución libre en <http://EZAnalyze.com>), o Excel. Aunque es muy común utilizar el programa SPSS para realizar una gran variedad de análisis estadísticos, su uso es relativamente complejo y especializado. Por ello, para el caso de aquellas personas que se encuentran dentro del área educativa y que no tienen muchos antecedentes en estadística, se recomienda el uso del programa Excel debido a que es accesible y su manejo resulta relativamente intuitivo, comparado con otros programas de cómputo más sofisticados.

Al utilizar el programa Excel para realizar el mencionado análisis de correlaciones, se requiere agregar un componente que usualmente no se encuentra instalado en las versiones básicas del programa. Dicho componente se denomina Análisis de Datos (Data Analysis) y se importa con relativa facilidad desde el menú de Herramientas (Tools); enseguida se selecciona la opción Complementos (Add-Ins) y a continuación se selecciona la opción Paquete de Herramientas de Análisis (Analysis ToolPack). De este modo, en el menú de Herramientas aparecerá disponible la opción Análisis de Datos (Data Analysis), misma que puede ser utilizada para estimar el índice de correlación entre la variable contextual analizada y el puntaje de los examinandos en la prueba, y que también indica el nivel de significancia estadística que tiene dicha correlación, lo cual aporta un elemento muy

importante para la interpretación de los resultados. Es importante mencionar que para obtener ambas estimaciones es necesario utilizar la opción Regresión (Regression) tras seleccionar Análisis de Datos (Data Analysis), ya que si se utiliza la función Correl incluida en la opción Función (Function) del menú Insertar (Insert), el programa solamente proporcionará la correlación entre las variables de interés, pero no incluirá el dato del nivel de significancia.

Ahora bien, la mayoría de los informes de resultados incluyen también otro tipo de información contextual respecto a variables existentes en la escuela o en el programa educativo, lo cual sirve para proveer un contexto más amplio para ayudar a la interpretación de resultados.

Estos datos contextuales pueden ser de utilidad para que el lector lleve a cabo una evaluación mucho más acertada de la ejecución del programa educativo en cuestión. Los informes de resultados generalmente incluyen información de variables como cuestiones demográficas de los estudiantes, datos socioeconómicos de la comunidad que los rodea, y algunos aspectos económicos relacionados con dicha comunidad; sin embargo, una aproximación más directa que apoya la interpretación de los resultados es la inclusión de un breve texto que contenga una explicación narrativa de lo que se encuentra en el informe de resultados.

Cabe destacar que, aunque la información contextual y descriptiva son elementos importantes de los informes de resultados, su función es meramente de apoyo, y deberá cuidarse que no opaquen en ningún momento los datos sobre los resultados de la evaluación, los cuales siempre deberán ser el foco central de un

informe de resultados. En este sentido, es importante mencionar que las investigaciones han demostrado que la mayoría de los destinatarios (incluyendo el público en general), buscan que el enfoque central de los informes sea precisamente los resultados; por ello si la información descriptiva es excesiva, la comprensión de los informes de resultados se tornará difícil (Condelli, 2005).

2. 3. Establecimiento de estándares de ejecución y puntos de corte

En el inciso 2.2, se comentó que comúnmente se presentan los porcentajes de aciertos obtenidos en la prueba en rangos que forman una escala ordinal, a fin de representar de manera cualitativa estándares de ejecución o niveles de dominio (por ejemplo, alto, medio y bajo) de los contenidos que miden los ítems del test. Para que puedan definirse tales estándares o niveles de dominio, se requiere establecer puntos de corte que delimiten los puntajes que separarán las categorías adyacentes que representan los distintos niveles de logro. La manera de hacerlo comúnmente involucra operaciones de juicio por parte de expertos. En esta sección se describirán diferentes procedimientos para establecer puntos de corte para las puntuaciones de pruebas criterioles.

Los puntos de corte en una prueba deben establecerse con base en los objetivos que persigue el instrumento, particularmente en cuanto al tipo de referencia que utilizamos para interpretar los resultados. Así, las pruebas que utilizan puntos de corte establecidos de manera normativa usualmente persiguen el objetivo de seleccionar a los estudiantes más aptos y asignar recursos a gran escala (Popham,

1990, citado por Contreras, 2000). Al establecer el punto de corte de manera normativa se utilizan porcentajes para interpretar los resultados de un estudiante en relación con las calificaciones de otros estudiantes, lo que constituye una interpretación relativa. Sin embargo, cuando se busca establecer puntos de corte para pruebas con propósito de certificación o de descripción de la ejecución en referencia al dominio de un criterio como el currículum, se busca diferenciar a los examinandos en dos categorías: los que tienen dominio del conocimiento evaluado y los que no lo tienen. Por ello es necesario establecer estándares de ejecución, los cuales se obtienen únicamente a través de la implementación de puntos de corte de tipo absoluto (Payne, 1992, citado por Márquez, 2005).

Así, en las pruebas criterios un punto de corte es una medida establecida en una escala de puntuación, utilizada para evaluar el desempeño de un examinando en relación a un dominio o competencia, y expresa el grado en que el examinando es o no competente para aplicar conocimientos, capacidades y habilidades adquiridas. En general, el punto de corte divide a los examinandos en dos grandes categorías: los que obtuvieron puntuaciones altas en la escala de calificación, llamados competentes; y los que obtuvieron una baja puntuación, denominados no competentes (Hambleton y Eignor, 1980; Jaeger, 1989; Popham, 1990).

Desde los años cincuenta del siglo pasado han existido diversos métodos para establecer puntos de corte (Angoff, 1971; Ebel, 1972; Nedelsky, 1954; citados por Márquez, 2005). Dichos métodos generalmente se basaban en juicios emitidos por

grupos de expertos, lo cual no resultaba totalmente confiable, ya que podía acreditar todo estudiante que demostrara tener conocimientos mínimos indispensables.

En el campo de la evaluación del aprendizaje existen dos modelos generales para establecer el punto de corte: los modelos estáticos y los continuos.

Los primeros consideran que el aprendizaje es dicotómico: se posee o no el dominio. Este modelo es funcional cuando el área o materia evaluada permite construir un examen con un elevado grado de homogeneidad en los reactivos, tal es el caso del área de las matemáticas; ya que el punto de corte se asume como el 100% de las respuestas correctas.

Los modelos continuos se basan en las teorías de adquisición del conocimiento, mismas que señalan que se van presentando grados o niveles de dominio conforme se da el aprendizaje.

Actualmente, algunos autores han buscado minimizar el factor de arbitrariedad del juicio de valor a través de la creación de más de 38 métodos, entre los cuales se pueden mencionar el de Berk, Hambleton, Livingston y Zieky, Jaeger y otros (Márquez, 2005). Berk, en 1986 propuso tres categorías para agrupar los procedimientos para determinar los puntos de corte: de jueceo, basado en opiniones que discuten grupos de expertos sin conocer el desempeño; de Jueceo-empírico, en donde los expertos sí tienen información acerca del desempeño; y el Empírico-Jueceo, donde se hace un análisis estadístico del desempeño de los examinandos y se define el criterio de dominio o no dominio. A continuación se encuentra una descripción de los modelos continuos que han causado mayor impacto a lo largo del

tiempo, y que han sido utilizados más ampliamente para establecer puntos de corte. En cuanto al resto de los métodos, se puede decir que muchos de ellos son modificaciones que surgieron a partir de los métodos clásicos que se describen aquí.

Método de Nedelsky (1954). Es un método de jueces que se utiliza con pruebas de opción múltiple. Se basa en el supuesto de que el examinando que posee una competencia mínima elimina en primer término las opciones de respuesta que a su juicio son incorrectas, y procede posteriormente a escoger la que considera correcta entre las que restan. Los jueces establecen qué opciones de cada ítem debería considerar erróneas el estudiante que tuviese la competencia mínima requerida para aprobar; posteriormente asignan el recíproco del número de opciones no seleccionadas por el examinando, de tal forma que si por ejemplo, en un ítem de cinco opciones los jueces consideran que un alumno mínimamente competente debería saber que son falsas tres de las cinco, el valor que se busca sería 0.5 (recíproco de dos). Este proceso genera tantos recíprocos como ítems. Posteriormente se obtiene la suma de todos los recíprocos, y el valor resultante se interpreta como la puntuación que obtendría el alumno mínimamente competente para el juez en cuestión. El punto de corte se establece como la media de los valores resultantes producidos por los jueces (Abad, F. J.; Atencia, J.; García, C.; Hontangas, P.; Olea, J.; Ponsoda, V.; Revuelta, J.; Suero, M.; y Ximénez, C., s/f). Entre las fallas de este método se encuentra el hecho de que no se puede saber con certeza si el examinando utiliza el método de adivinación o no, además de que los

valores asignados por los jueces regularmente quedan alrededor de .5, lo cual produce un punto de corte más bajo que otros métodos.

Método de Angoff (1971). También es un método de jueceo donde el grupo de especialistas revisa los resultados y estima la probabilidad de que un examinando pueda contestar correctamente la pregunta. Se suman las estimaciones de los especialistas y se obtiene el promedio de ellas, mismo que se utilizará para establecer el punto de corte. Es un método recomendado cuando se persiguen fines de certificación, debido a que permite tomar decisiones en una línea continua de probabilidades, lo cual produce puntos de corte más directos y verosímiles; sin embargo, generalmente se obtienen variaciones grandes entre las probabilidades en cada reactivo de cada juez, lo cual presenta una desventaja.

Método de Ebel (1972). Método de jueces donde se analiza la dificultad y relevancia de cada reactivo. Implica dos etapas: cada juez clasifica primero las preguntas en grupos y luego emite un solo juicio numérico para cada grupo de preguntas. La clasificación de preguntas en grupos se basa en dos tipos de juicios sobre cada pregunta: sobre su dificultad y sobre su relevancia. Ebel sugirió tres niveles de dificultad, denominados “fácil,” “mediano” y “difícil”; y cuatro categorías de importancia, denominadas “esencial,” “importante,” “aceptable” y “objetable.” El juez primero clasifica todas las preguntas del examen. Después emite juicios sobre el rendimiento de un examinando que posee una competencia mínima. El juez debe emitir un juicio por cada uno de los 12 bloques de la tabla de clasificación (excepto los que estén vacíos). Es decir, debe emitir un juicio para las preguntas clasificadas

“esenciales, fáciles”; otro para las preguntas clasificadas “esenciales, medianas”; y así sucesivamente hasta llegar a “objetable, difícil.” El juicio consiste en una respuesta a la pregunta: “Si un examinando con competencia mínima respondiera muchas preguntas como éstas, ¿Qué porcentaje contestaría correctamente?”. Considerar al mismo tiempo la dificultad y la importancia de cada pregunta es especialmente valioso cuando las preguntas del examen difieren ampliamente en importancia. Sus desventajas son su lentitud y su inadecuación para exámenes cortos (Livingston y Zieky, 2004).

Método de Jaeger (1978). Este método está clasificado en la categoría de Jueceo-Empírico y se considera interactivo porque los jueces que participan tienen diversos marcos de referencia y utilizan información normativa del desempeño de los examinandos. En lugar de buscar competencias mínimas, se busca si el examinando puede o no contestar correctamente el reactivo, y con ello determinar si se le otorga o no el documento que acredite su conocimiento. El punto de corte es el cálculo del promedio de las respuestas de los jueces. El proceso interactivo de este método permite reducir la variabilidad en la distribución de los estándares recomendados, y con ello incrementar la confiabilidad del estándar. Una desventaja es que consume mucho tiempo y es más complejo que otros métodos de jueceo.

Método de Livingston y Zieky (1982). Es un método del tipo Empírico-Jueceo, también conocido como *Método de contraste de grupos*. En lugar de enfocarse en el examen, lo hace en los examinandos a quienes clasifica en dos grupos: aptos o No Aptos. Después se comparan las distribuciones de las

puntuaciones de ambos grupos y el punto de corte es donde se intersectan ambas distribuciones. Este método es aparentemente simple, pero el punto de corte puede verse afectado por factores como la proporción de los no calificados en la muestra, la severidad con que los jueces definen el dominio, la decisión de clasificar los casos dudosos (eliminarlos, considerarlos aptos, o no aptos).

En la tabla 3 se presenta una síntesis de los métodos descritos para establecer puntos de corte para las pruebas criterioles.

Tabla 3. Modelos continuos más comunes para establecer puntos de corte.

Procedimiento	Modelo	Características	Ventajas/Desventajas
Jueceo	Nedelsky (1954)	▪ Utilizado solo con pruebas de opción múltiple. ▪ Basado en el supuesto de que el examinando elimina en primer término las respuestas que a su juicio son incorrectas, y procede posteriormente a escoger la que considera correcta entre las que restan. Forma de obtención del PC ▪ Los jueces establecen qué opciones del ítem debería considerar erróneas el estudiante con competencia mínima; ▪ Asignan el recíproco del número de opciones no seleccionadas por el examinando; p. ej: en un ítem de 5 opciones los jueces consideran que un alumno mínimamente competente debería saber que son falsas 3 de las 5, el valor que se busca sería 0.5 (recíproco de 2). ▪ Se obtiene la suma de todos los recíprocos. ▪ El valor resultante será la puntuación del alumno mínimamente competente para el juez en cuestión. ▪ El punto de corte es la media de los valores resultantes producidos por los jueces.	Desventajas: ▪ No se puede saber con certeza si el examinando utiliza el método de adivinación o no. ▪ Los valores asignados por los jueces regularmente son .5, lo cual produce un punto de corte más bajo que otros métodos. ▪ Requiere mayor entrenamiento de los expertos, provocando que parezca ser más científico que otros métodos
	Angoff (1971)	▪ El grupo de especialistas revisa los resultados y estima la probabilidad de que un examinando pueda contestar correctamente la pregunta. Forma de obtención del PC: ▪ Se suman las estimaciones de los especialistas ▪ Se determina el promedio de las estimaciones ▪ Se utiliza el promedio para establecer el punto de corte.	Ventajas: ▪ Fácil de entender, implementar y calcular. ▪ El punto de corte es específico a las preguntas de la prueba y juicios de valor de los especialistas ▪ Método recomendado cuando se persiguen fines de certificación porque permite tomar decisiones en una línea continua de probabilidades, lo cual produce puntos de corte más directos y verosímiles. Desventajas: ▪ Generalmente se obtienen variaciones grandes entre las probabilidades de los reactivos de cada juez
	Ebel (1972)	▪ Se hace un análisis de la dificultad y relevancia de cada reactivo. Forma de obtención del PC: ▪ Implica un procedimiento de dos etapas ▪ Se estima el porcentaje de reactivos de cada celda que puede contestar correctamente un examinando. ▪ Para obtener el promedio se multiplica el número de reactivos de las celdas por el porcentaje respectivo. ▪ Se suman todos los resultados y divide entre el número de reactivos (promedio ponderado) 1.- Cada juez clasifica las preguntas en grupos, basándose en su dificultad (fácil, mediano y difícil), y su relevancia (esencial, importante, aceptable y objetable) 2.- Emiten juicios por cada uno de los 12 bloques de la tabla de clasificación, sobre el rendimiento de un examinando con competencia mínima.	Ventajas: ▪ Es un método muy valioso cuando las preguntas del examen difieren ampliamente en importancia. Desventajas: ▪ Es un método inadecuado para exámenes cortos ▪ Su procedimiento implica mucho tiempo.
Jueceo-Empírico	Jaeger (1978)	▪ Clasificado como interactivo porque los jueces que participan tienen diversas características (maestros, administradores, etc.), y porque utiliza información normativa del desempeño de los examinandos. ▪ No busca competencias mínimas, busca si el examinando puede o no contestar correctamente el reactivo ▪ Así se determina si se le otorga o no el documento que acredite su conocimiento. Forma de obtención del PC: ▪ Es el cálculo del promedio de las respuestas de los jueces.	Ventajas: ▪ El proceso interactivo de este método permite reducir la variabilidad en la distribución de los estándares recomendados, y con ello incrementar la confiabilidad del estándar. Desventajas: ▪ Consume mucho tiempo y es más complejo que otros métodos de jueceo.
Empírico-Jueceo	Livingston y Zieky (1982)	▪ También conocido como Método de contraste de grupos. ▪ Se basa primordialmente en los examinandos y no en el examen. Se clasifica a los examinandos en uno de dos grupos: aptos o no aptos. Forma de obtención del PC: ▪ Se combinan las distribuciones de las puntuaciones de ambos grupos ▪ El punto en donde se intersectan ambas distribuciones es el punto de corte.	Ventajas: ▪ Es un método sencillo y simple Desventajas: ▪ El punto de corte puede verse afectado por factores como ○ La proporción de los no calificados en la muestra, ○ La severidad con la que los jueces definen el dominio; ○ La decisión que se toma para clasificar los casos dudosos (eliminarlos, aptos, no aptos).

UNIDAD 3. Procedimiento para elaborar informes de resultados.

Objetivo particular: los participantes conocerán los elementos básicos para generar informes técnicos de los resultados en el examen, según las necesidades de los destinatarios.

Introducción a la Unidad

Hasta el momento, se ha contextualizado la participación del CARIRE en el proceso de desarrollo de la prueba y se han descrito los procedimientos necesarios para efectuar diferentes tipos de análisis de la información derivada de la aplicación de la prueba a gran escala. A continuación se describirán los procedimientos necesarios para elaborar informes de resultados efectivos que estén acordes con las necesidades informativas de los distintos usuarios de la información.

3.1. Desarrollo de informes de resultados

En esta sección se describen los propósitos y funciones de los informes dentro del proceso de construcción de los exámenes, así como los procedimientos que se seguirán para desarrollarlos y los productos esperados.

Para elaborar informes de resultados efectivos acordes con las necesidades de los usuarios de la información, deberán tomarse en cuenta los siguientes aspectos:

- Determinar la forma en que los resultados individuales y grupales ayudarán a las escuelas en el mejoramiento del proceso educativo.
- Por otro lado, el informe deberá explicar cuáles fueron los propósitos que orientaron la desagregación de la información, quiénes son los destinatarios de la misma, qué información debe proporcionárseles y qué usos podrán darle. Como ya se comentó, los principales destinatarios incluyen a los mismos estudiantes, profesores, autoridades educativas, padres de familia y público en general, entre otros. Deberá efectuarse un análisis fundado, respecto a las razones por las cuales se deben reportar determinados resultados de la evaluación a estos destinatarios en particular.
- De igual manera, deberá cuidarse la claridad respecto a la información que se comunique en el informe, sus propósitos y técnicas utilizadas al reportarla (Roeber *et al*, 1980; citado por Roeber, 1995).
- Describir el contexto dentro del cual se llevó a cabo la evaluación; es decir, mencionar las características del centro educativo que lo distinguen de otros, y que a su vez describan de manera detallada sus particularidades, tales como: nombre de la escuela, descripción socioeconómica del entorno en que se ubica, cantidad de alumnos que se atienden, características socioeconómicas y culturales del alumnado, grados académicos que se imparten, materias o áreas que se evaluaron; entre otras cosas.
- Además se debe describir la naturaleza de la evaluación (prueba). Este aspecto del informe debe contemplar toda la información que contextualice al usuario respecto

al tipo de examen que se aplicó y la forma en que fue elaborado; desde su planeación, construcción, aplicación y análisis de resultados.

- Finalmente, todo informe de resultados debe incluir gráficas, tablas, esquemas y otros elementos que describan de forma clara, abreviada y significativa la información obtenida a partir del análisis de los resultados de la ejecución de los examinandos en la prueba. Para ilustrar este punto, en la figura 3 se puede observar un ejemplo sobre cómo elaborar gráficas con esas características a partir del análisis de resultados.

ESPAÑOL, COMPRENSIÓN DE LECTURA Y REFLEXIÓN SOBRE LA LENGUA

Figura 2. Porcentaje de estudiantes por nivel de logro educativo y estrato escolar: Español/ Comprensión de la lectura y Reflexión sobre la lengua

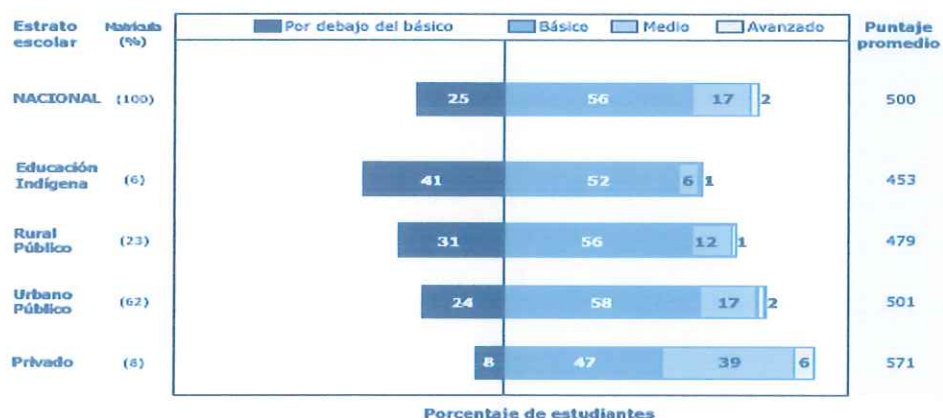


Figura 3. Ejemplo de cómo elaborar gráficas a partir del análisis de resultados de un examen criterial de gran escala

Fuente: Martínez, F. (2007). El aprendizaje en tercero de primaria en México: español, matemáticas, ciencias naturales y ciencias sociales. Resumen ejecutivo basado en: Backhoff, E. Andrade, E., Sánchez, A. y Peón, M. (2007). El Aprendizaje en Tercero de Primaria en México: Español, Matemáticas, Ciencias Naturales y Ciencias Sociales. México: INEE.

Usualmente, en el informe de resultados una gráfica como la que aparece en la figura 3 debe ser seguida por un comentario aclaratorio como el siguiente:

se puede observar en la figura que "...los cuatro segmentos de cada barra muestran la proporción de alumnos que se ubica en cada uno de los niveles de competencia definidos; las gráficas tienen una barra para cada servicio educativo de la muestra... para dimensionar la importancia de los resultados, según el tamaño de las poblaciones correspondientes, al lado del nombre de cada estrato, entre paréntesis, se anota el porcentaje de la matrícula nacional que atienden las escuelas de cada uno; a la derecha de cada barra se anota el puntaje promedio de los alumnos de cada servicio educativo... en la figura se puede apreciar que, en el área de español y a nivel nacional, uno de cada cuatro estudiantes (25%) se encuentra en el nivel por debajo del básico; 56% se ubica en el nivel básico; 17% se sitúa en el nivel medio y sólo dos alumnos de cada cien alcanzan el nivel avanzado. Las diferencias que distinguen a los alumnos de los servicios educativos se aprecian considerando que por debajo del nivel básico se encuentran, en orden descendente, 41% de los alumnos de educación indígena; 31% de niños de las escuelas rurales públicas; 24% de los estudiantes de escuelas urbanas públicas; y, finalmente, 8% de quienes acuden a las escuelas privadas" (Martínez, 2007, pp. 29-30).

3.2. Tipos de destinatarios de la información

Esta sección complementa la información presentada en los incisos 1.2.2, 2.2, y 3.1 sobre los destinatarios de los informes del examen.

Existen diversos procedimientos que se pueden utilizar cuando se elaboran informes de resultados de un examen de gran escala, y la elección de ellos depende

críticamente del tipo de destinatario al cual está dirigida la información. Por su parte, el aspecto más importante sobre el cual versará el informe es el tipo de información que resulta más relevante para apoyar la toma de decisiones en el nivel correspondiente a cada destinatario. A continuación se presentan los principales usuarios de la información del examen y los principales datos que son de su incumbencia:

- **Estudiantes:** al reportarles sus resultados se recomienda hacerlo en dos fases: la primera implica proporcionar un informe general de la ejecución de los examinandos que ya recibieron sus resultados individuales. La segunda se trata de sesiones de seguimiento individual con cada estudiante, enfocadas sobre la forma en que los maestros abordarán las necesidades individuales de los estudiantes.
- **Padres de familia:** dado el interés para los padres de familia por conocer el rendimiento académico de sus hijos, tanto en forma individual como en relación a todo el estudiantado y otras escuelas, los informes deben incluir estos aspectos como punto medular. Para ello se sugieren cuatro estrategias:
 1. Reuniones individuales entre el maestro y los padres de familia. No obstante, aunque es lo más recomendable, debido a situaciones particulares no siempre es posible llevarlas a cabo.
 2. Informes individuales enviados a casa. Pueden proporcionar una forma de comunicación accesible maestro – padre de familia, pero deberán incluir qué tan buena o mala fue la ejecución del examinando, así como los pasos a seguir después de la evaluación, que incluyen la forma en que los padres

pueden participar en un plan de acción para apoyar la educación de sus hijos.

3. Reuniones con padres de familia.

4. Boletines informativos dirigidos específicamente a los padres de familia.

- **Funcionario de primer nivel (Secretario de Educación, Rector, Director General, etc.)** Debido a que se trata de entidades que se encargan del establecimiento de políticas educativas a nivel estatal o institucional, el tipo de información reportada debe incluir en primer lugar, una distribución de promedios o porcentajes de aciertos de los planteles del estado, municipio, campus, región, institución, etc., por cada área curricular. En segundo término, se debe reportar la tendencia del municipio, campus, región, etc., y los promedios en cada eje curricular, durante los cinco años anteriores. Por último, deberá incluirse en forma gráfica, una comparación simultánea de los resultados de logro vs. las acciones educativas determinadas (perfil de profesores, recursos de la escuela, etc.). Para ello se sugiere una estrategia que implica tres fases:

En la primera fase se debe elaborar un informe de los antecedentes que rodean la evaluación, que incluya una explicación de lo que fue evaluado, los tipos de evaluación empleados y motivos por los cuales se eligieron, así como la forma en que se utilizarán y reportarán los resultados. Este informe deberá ser entregado en el momento en el que se esté recolectando la información, y antes de que los resultados de la evaluación estén disponibles. De esta forma, se invita a este tipo de

destinatario a enfocarse en el trasfondo de la evaluación, y no únicamente en los datos cuantitativos.

En la segunda fase, se incluirá un informe que contenga los resultados de la evaluación nacional, estatal, municipal, etc., y que responda a preguntas tales como:

- ¿Qué resultados se obtuvieron respecto a la ejecución a nivel nacional, estatal, municipal, etc.?
- ¿Cómo se presentan los resultados en comparación al año anterior?
- ¿En qué nivel se encuentran los puntajes de los examinandos: básico, medio, alto?
- ¿Qué medidas se están tomando para examinar los resultados de la evaluación y mejorar el proceso educativo?
- ¿Cómo se elaborarán los informes que serán dirigidos a los padres de familia y público en general?

El tercer informe debe proporcionar un seguimiento de los esfuerzos realizados para mejorar el proceso educativo en el estado o, en su caso, el proceso instruccional en la escuela, así como de su efectividad. Aunque se propone como un informe opcional, es muy efectivo para orientar a este tipo de destinatario hacia el verdadero propósito de la evaluación, el cual es mejorar el proceso educativo, y no solamente proporcionar puntajes para establecer niveles de calidad en las escuelas.

- **Directores de planeación educativa, directores académicos y otros similares.**

Los datos más relevantes incluyen: promedios de ejecución de estudiantes en cada

eje curricular y en grupos de preguntas que evalúan blancos de aprendizaje claves por eje curricular; resultados en ambos por campus, región, municipio, etc.

- **Inspector, coordinador, etc.** La información que resulta relevante en este nivel incluye los promedios de resultados por plantel dentro del ámbito o región de servicio, por cada eje curricular; así como la tendencia de la ejecución promedio, durante los cinco años anteriores.
- **Director de plantel.** El tipo de información relevante que debe reportarse es más específico, ya que lo que interesa es el comportamiento de los examinandos dentro de una escuela específica. Incluye resultados de los examinandos del plantel en cuanto a promedio general y por género; porcentaje de aciertos de cada examinando; comparación del promedio general y por género, de los resultados de los examinandos del plantel, contra las correspondientes medias de los planteles del campus, región, municipio, etc. Todo lo anterior por eje, subeje curricular ó línea de formación que define el currículum (Contreras y Backhoff, 2004). Debido a que este tipo de destinatario tiene relación directa con la implementación de estrategias y acciones encaminadas al mejoramiento académico del alumno, es importante incluir en el informe elementos que proporcionen directrices para llevar a cabo sesiones de seguimiento individual con cada estudiante, enfocadas en la forma en que los maestros abordarán las necesidades individuales de los mismos.
- **Público en general:** lo primero que debe hacerse es decidir el tipo de público al cual se va a dirigir la información, ya que es un grupo muy heterogéneo que tiene distintas necesidades de información y conocimiento. Enseguida se deben explicar

las metas y propósitos del informe. Por último, se deben establecer los medios por los cuales se dará a conocer la información.

La tabla 4 resume la información de los principales destinatarios de los informes de resultados del examen aquí presentada.

En conclusión, la elaboración de informes de resultados a partir de los análisis de la ejecución de los examinandos, no solo será un insumo fundamental para los evaluadores educativos, directivos de primer nivel, directores de plantel y demás instancias educativas, sino que al darse a conocer tendrán un importante impacto en las comunidades académicas de las instituciones participantes. Por ello, cada informe deberá ser elaborado de conformidad con los requerimientos informativos de cada destinatario, así como los estándares de calidad técnica existentes, de manera que resulte riguroso desde el punto de vista del contenido y del psicométrico.

INSERTAR AQUÍ **Tabla 4**

Destinatarios de informes.doc

(Archivo independiente llamado igual).

Tabla 4. Destinatarios de los informes de resultados del examen

Tipo de organización	Tipo de examen	Destinatario	Tipo de información a ser reportada
Secretaría estatal de educación	Examen estatal	Primer nivel: Secretario de Educación del Estado y Subsecretario de Educación Básica	1. Distribución de promedios de las escuelas del estado, en cada eje curricular 2. Distribución de promedios de las escuelas, por municipio, en cada eje curricular 3. Tendencia del municipio y promedios estatales en cada eje curricular, durante los cinco años anteriores 4. Gráfica que compara simultáneamente los resultados de logro contra acciones educativas determinadas (perfil de profesores, recursos de la escuela, etc.)
		Segundo nivel: Director general de planeación educativa	1. Promedio de ejecución de estudiantes en grupos de preguntas que evalúan blancos de aprendizaje claves, por eje curricular 2. Promedio de ejecución de estudiantes en cada eje curricular 3. Resultados en 1 y 2, por municipio
		Tercer nivel: Inspector	1. Promedio de resultados por escuela dentro de la región de servicio, para cada eje curricular 2. Tendencia de la ejecución promedio, durante los cinco años anteriores
		Cuarto nivel: Director de Escuela, subdirector y profesor	1. Promedio general y por género, de resultados de los estudiantes de la escuela, por eje curricular 2. Porcentaje de aciertos por eje curricular, para cada examinando 3. Comparación del promedio general y por género, de los resultados de los estudiantes de la escuela, por eje curricular, contra las correspondientes medias de las escuelas del estado
		Otros niveles: Estudiante y padre	1. Porcentaje de aciertos por eje curricular, obtenidos por el estudiante 2. Comparación del porcentaje promedio de aciertos del grupo, por eje curricular, contra el correspondiente logrado por el estudiante
Institución educativa estatal	Examen estatal	Primer nivel: Director general	1. Distribución de promedios de las escuelas del estado, en cada unidad 2. Distribución de promedios de las escuelas, por municipio, en cada unidad 3. Tendencia del municipio y promedios estatales en cada unidad, durante los cinco años anteriores 4. Gráfica que compara simultáneamente los resultados de logro contra acciones educativas determinadas (perfil de profesores, recursos de la escuela, etc.) 5. Porcentaje de aciertos de las escuelas del estado, por municipio, en cada unidad
		Segundo nivel: Director de planeación académica	1. Promedio de ejecución de estudiantes en grupos de preguntas que evalúan blancos de aprendizaje claves, por unidad y tema 2. Promedio de ejecución de estudiantes en cada unidad y tema 3. Resultados en 1 y 2, por municipio 4. Porcentaje de aciertos de los estudiantes, en cada unidad y tema evaluados
		Tercer nivel: Director de plantel	1. Promedio de resultados por grupo, para cada unidad y tema 2. Tendencia de la ejecución promedio, durante los cinco años anteriores
		Cuarto nivel: Jefe de materia y profesor	1. Promedio general y por género, de resultados de los estudiantes de la escuela, por unidad y tema 2. Porcentaje de aciertos por unidad y tema, para cada examinando 3. Comparación del promedio general y por género, de los resultados de los estudiantes de la escuela, por unidad y tema, contra las correspondientes medias de las escuelas del estado
		Otros niveles: Estudiante	1. Porcentaje de aciertos por unidad y tema, obtenidos por el estudiante 2. Comparación del porcentaje promedio de aciertos del grupo, por unidad y tema, contra los correspondientes logrados por el estudiante
Escuela	Examen departamental o colegiado	Primer nivel: Director	1. Distribución de promedios de las carreras de la escuela, en cada unidad 2. Tendencia de la escuela y promedios en cada carrera, durante los cinco años anteriores 3. Gráfica que compara simultáneamente los resultados de logro contra acciones educativas determinadas (perfil de profesores, recursos de la escuela, etc.)
		Segundo nivel: Subdirector	1. Promedio de ejecución de estudiantes en grupos de preguntas que evalúan blancos de aprendizaje claves, por unidad y tema 2. Promedio de ejecución de estudiantes en cada unidad y tema 3. Resultados en 1 y 2, por carrera
		Tercer nivel: Coordinador de carrera	1. Promedio de resultados por grupo, para cada unidad y tema 2. Tendencia de la ejecución promedio, durante los cinco años anteriores
		Cuarto nivel: Jefe de materia y profesor	1. Promedio general y por género, de resultados de los estudiantes de la escuela, por unidad y tema 2. Porcentaje de aciertos por unidad y tema, para cada examinando 3. Comparación del promedio general y por género, de los resultados de los estudiantes de la carrera, por unidad y tema, contra las correspondientes medias de las carreras de la escuela
		Otros niveles: Estudiante	1. Porcentaje de aciertos por unidad y tema, obtenidos por el estudiante 2. Comparación del porcentaje promedio de aciertos del grupo, por unidad y tema, contra los correspondientes logrados por el estudiante

3. 3. Análisis, reporte e interpretación de resultados en forma desagregada

Actualmente se utilizan los resultados de las pruebas criteriosales en áreas muy diversas. Por ejemplo, los maestros, utilizan los resultados de la ejecución de los alumnos para ubicarlos en programas educativos adecuados a sus necesidades, monitorear sus avances, realimentar su aprendizaje, identificar sus deficiencias y también para planear su instrucción. En el área de educación especial este tipo de informe ayuda mucho en el diagnóstico de problemas de aprendizaje, así como en el monitoreo del progreso de los alumnos. También se utiliza este tipo de informes para evaluar programas educativos normales y avanzados. Pero quizás uno de los usos más importantes que se le da a los resultados de las pruebas se ubica en el área de certificación profesional (Hambleton, 1988).

En el ámbito internacional, en 1994 el Congreso de los Estados Unidos aprobó el Acta de las Escuelas de América documento en el cual se estipula en el Título I, como una obligación, y a la vez como una oportunidad para los estados, organismos educativos y escuelas, desagregar y analizar los datos resultantes de las evaluaciones de sus estudiantes (Jaeger y Tucker, 1998). Estas evaluaciones permitirán la desagregación de la información en variables específicas, tales como género, escuela, desempeño en el área de matemáticas, estatus migratorio, discapacidades, situación económica, etc. Sin embargo, esto implica una necesidad de definir adecuadamente dichas variables. Entre otras cosas, es necesario especificar definiciones de las categorías de los estudiantes, de los aspectos legales en cuanto a la confidencialidad de los resultados, de la generalización adecuada de

los resultados de la evaluación, de la interpretación válida de los resultados de la evaluación dentro y entre las categorías de los estudiantes, y de la difusión de los resultados de tal forma que sean útiles a distintos destinatarios.

Lo anterior brinda una gran oportunidad para llevar a cabo un análisis nuevo y creativo de la información, cuyo uso podría extenderse al análisis de las políticas educativas actuales, evaluación de los programas existentes, planeación y desarrollo de intervenciones instruccionales innovadoras, así como la elaboración de informes respecto al estado actual y al progreso de los alumnos. Quienes formen parte activa del sistema educativo y tengan injerencia en el análisis y reporte del desempeño colectivo de los alumnos, se verán beneficiados al desagregar la información que arrojan los resultados de las evaluaciones. Dentro de esa población se encuentran maestros, directivos, inspectores de zona, administradores, instituciones de educación y personal del área de la evaluación educativa, e inclusive agencias gubernamentales que toman decisiones en torno a políticas educativas a nivel local, estatal y/o nacional (Jaeger y Tucker, 1998).

Jaeger y Tucker (1998) proponen tres escenarios para elaborar informes de resultados: la Zona escolar, la Escuela y el Estado. Los autores explican cuáles son los elementos y procedimientos que se deben incluir en todo informe, y la importancia de reportar en forma desagregada dicha información.

Zona escolar

En el caso de la zona escolar, es necesario abordar los siguientes aspectos:

1. La forma en que se desagregarán los resultados de la prueba para entender de mejor manera los factores asociados al desempeño de los estudiantes.
2. La manera en que se presentarán dichos resultados, de tal forma que se resalten los aspectos más interesantes de la ejecución de los alumnos.
3. Si el tamaño del grupo de estudiantes evaluados es lo suficientemente grande como para respaldar el informe.
4. Si las diferencias entre las ejecuciones promedio de los estudiantes de varios grupos es lo suficientemente confiable.
5. El tipo de conclusiones que se pueden obtener respecto a las causas de los resultados obtenidos.

De esta manera se podrán obtener informes de resultados cuyas principales funciones sean proporcionar información útil para que los directores de planteles y maestros de una zona escolar específica puedan llevar a cabo una adecuada planeación curricular, para que las autoridades educativas utilicen la información cuando se realicen revisiones de las políticas educativas, y también para que se difundan los resultados entre la población que conforma dicha zona escolar, así como entre los padres de familia cuyos hijos se encuentran estudiando dentro de ella.

Para que esto pueda llevarse a cabo, se requiere analizar los resultados de cada zona escolar que se encuentra en el municipio y en el estado, y los de cada escuela en cada zona, en forma separada.

Se debe examinar a los alumnos por lo menos una vez en el nivel escolar que se esté evaluando (básico, medio, medio superior o superior); y los resultados de la evaluación se deben analizar y reportar por zona y escuela en forma separada, ubicando por categorías a los estudiantes (género, nivel socioeconómico, turno, etc.). Para que esta información sea útil para los maestros cuando lleven a cabo su planeación curricular, se sugiere que:

1. Se resuman los resultados de la evaluación de los alumnos de cada grupo.
2. Una parte del resumen describa la ejecución general del grupo en lo que respecta a la materia evaluada.
3. Otro tipo de información que debe incluir el informe es si la ejecución de los alumnos fue similar, o si algunos alumnos tuvieron mejores resultados que otros.
4. Es muy importante que al reportar los resultados se especifique la ejecución de los alumnos en relación a los reactivos que evalúan dominios específicos del conocimiento general que evalúa la prueba.
5. También se debe incluir un marco de referencia para que el docente sepa discriminar entre los puntajes que distinguen una buena ejecución de una excelente; y, si se encuentra disponible, establecer una comparación entre puntajes actuales con puntajes de un año anterior.

En el caso de los Directores de plantel, para que la información de los resultados de la evaluación resulte útil en la planeación curricular, se sugiere que:

1. Se genere un documento que sintetice los resultados de los estudiantes en los momentos académicos críticos, según el nivel educativo que estén cursando, de tal manera que cada Director conozca los resultados de su plantel y de los otros que conforman la zona, para que pueda establecer puntos de comparación.
2. Se disponga la información para que se pueda establecer una comparación entre la ejecución de años anteriores y el actual.
3. Al reportar la información se pueda observar si existen diferencias entre los puntajes de los examinandos, con la finalidad de saber si la ejecución fue similar para todos o no.
4. Se establezcan en el informe las categorías de estudiantes cuyo porcentaje de eficiencia se considere pobre, regular, bueno y excelente (o la escala correspondiente). Al respecto, véase el inciso 2.3 sobre el establecimiento de estándares y puntos de corte.
5. Se agrupe la información por variables de contexto, tales como género, turno, nivel socioeconómico, etc.

Existen cuatro tipos de análisis que se pueden producir con los datos: el primero es una categorización de los porcentajes obtenidos con las puntuaciones de los examinandos, agrupándolos por variable de contexto (edad, género, turno, etc.), y según la escala que se esté utilizando (p. ej. excelente, bueno, regular, malo). El segundo es la obtención de los puntajes promedio, por variable. El tercer análisis implica el uso de gráficas, donde una de las más útiles es la de cajas y bigotes (ver

figura 4) que presenta la información en rangos percentiles, ya que facilita la comparación de la distribución de los puntajes de diversos grupos. Por último, se puede hacer un análisis estadístico de la variación de los puntajes promedio generada por un muestreo aleatorio y por una prueba poco confiable.

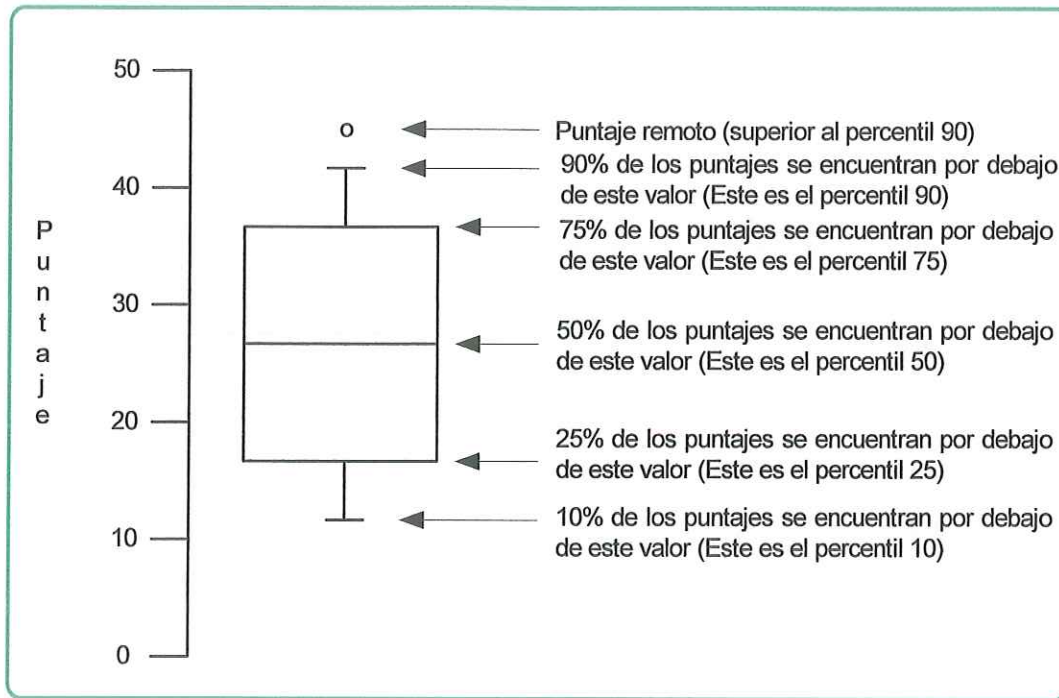


Figura 4. Ejemplo de gráfica de cajas y bigotes.

Fuente: Jaeger, R. M. y Tucker, C. G. (1998). Analyzing, disaggregating, reporting, and interpreting student's achievement test results: A guide to practice for Title I and beyond. Estados Unidos: Council of Chief State School Officers.

Escuela

En el caso del informe a una escuela, es necesario plantear una serie de preguntas que servirán como guía al realizar un análisis de datos:

1. ¿De qué manera se pueden desagregar los resultados para tener una mejor comprensión de los factores asociados con los avances colectivos de los estudiantes en un periodo de dos años?
2. ¿Cómo se puede llevar a cabo una comparación de los resultados de la ejecución de los alumnos en un periodo de dos años, de tal forma que refleje los avances de toda la escuela en relación a las metas académicas definidas?
3. ¿Qué tan grande debe ser el grupo de alumnos como para apoyar el informe de su ejecución global en el examen?
4. ¿Qué tan confiables son las diferencias entre los promedios de la ejecución de los estudiantes en varios años consecutivos?

Para poder desarrollar una estrategia de análisis de datos, deben revisarse aspectos tales como la cantidad de alumnos evaluados en los diferentes ciclos escolares, las variables de contexto de los examinandos, y la confiabilidad de los datos obtenidos. En este sentido, es recomendable realizar el siguiente análisis estadístico:

- Distribución de frecuencias de los examinandos, por ciclo escolar y variables de contexto de las cuales se tenga información (edad, grupo, género, nivel socioeconómico, turno, etc.)
- Puntajes promedio y desviaciones estándar de todos los examinandos, de los examinandos agrupados por variables de contexto, y de cada ciclo escolar.

- Distribución de los porcentajes de estudiantes ubicados en cada nivel de dominio (previamente determinado), por ciclo escolar y por variables de contexto-ciclo escolar.
- Intervalos de confianza de los puntajes promedio globales de cada ciclo escolar evaluado, así como por variables de contexto; o una prueba de hipótesis de las diferencias entre los puntajes promedio de los ciclos escolares que se deseen comparar.

Una vez que se lleve a cabo el análisis estadístico, se debe proceder a graficar los resultados, de tal manera que éstos se puedan apreciar en forma abreviada y clara. Se recomienda utilizar gráficas sencillas (barras, columnas, etc.). Como resultado final se pueden obtener propuestas de estrategias de mejoramiento del currículum, tales como revisiones exhaustivas de los contenidos de los cursos, materias o conocimientos que se evalúan, en contraste con el currículum propuesto, con la finalidad de saber si los estudiantes están aprendiendo lo que se supone; ó la revisión de la ejecución de los alumnos en exámenes parciales para determinar si existe la necesidad de crear programas remediales que aborden contenidos específicos en donde los alumnos estén teniendo dificultades.

Estado

El tipo de cuestionamientos que se deben abordar para realizar el análisis de datos y elaboración de informes cuando se tiene información de evaluaciones a nivel estatal, son:

1. ¿Qué tipo de análisis se pueden llevar a cabo con la finalidad de vislumbrar las tendencias de ejecución de los estudiantes en una determinada cantidad de ciclos escolares?
2. ¿De qué manera se pueden desagregar los resultados para tener una mejor comprensión de los factores asociados con los avances globales de los estudiantes en una determinada cantidad de ciclos escolares?
3. ¿Cuándo resultan estadísticamente confiables y académicamente significativos los promedios de ejecución de los estudiantes en evaluaciones de ciclos escolares consecutivos?
4. ¿Cuáles son los factores que pueden contribuir año tras año a las tendencias de la ejecución de los examinandos?
5. ¿De qué manera se puede modelar la tendencia de la ejecución global de los examinandos?

Las personas que toman decisiones y llevan a cabo políticas educativas requieren informes de resultados que reflejen tendencias de ejecución de los estudiantes a través de diferentes ciclos escolares, de tal manera que puedan determinar si existen evidencias de cambios sustanciales a través del tiempo, o evidencia de diferencias entre la ejecución evaluada y la ejecución por grupos o variables de contexto (género, turno, escuela, nivel socioeconómico, escolaridad de los padres, etc.). En pocas palabras, a esta población le interesa saber quiénes son los que cuentan con las habilidades que establecidas por el currículum, y quiénes no.

Para elaborar este tipo de informes, se debe recopilar información de evaluaciones realizadas durante varios años consecutivos, así como realizar los análisis estadísticos y gráficos necesarios. Cabe destacar que en la medida en que se desagrega la información, se debe definir con mayor claridad a los subgrupos de los cuales se reporta la información, ya que de no hacerlo, se corre el riesgo de generar un informe, así como una interpretación de los resultados equivocados.

En conclusión, la función principal de los informes que desagregan la información recabada de una evaluación a gran escala es hacer visibles los avances (o retrasos) de diversos subgrupos de estudiantes dentro de un sistema educativo, de tal manera que brinden una oportunidad para dirigir adecuadamente los recursos con los que se cuentan dentro del campo de la educación. Dichos informes deben abordar los siguientes aspectos:

- ✓ Definiciones de las categorías para la desagregación
- ✓ Tamaño mínimo de los grupos con los cuales se reportarán los resultados
- ✓ Error de medida y de muestreo
- ✓ Construcción de los intervalos de confiabilidad
- ✓ Métodos para mostrar la información (diferentes gráficas y tablas)
- ✓ Complejidad para determinar la causa de los patrones específicos de aprovechamiento
- ✓ Interpretación adecuada de los datos grupales
- ✓ Interpretación adecuada entre grupos, a través del tiempo

Todos estos aspectos, aproximaciones para analizar datos y demás recomendaciones se pueden generalizar a cualquier contexto en el cual se requiera interpretar los resultados de la ejecución de examinandos en una prueba, con la finalidad de proporcionar información para la toma de decisiones educativas.

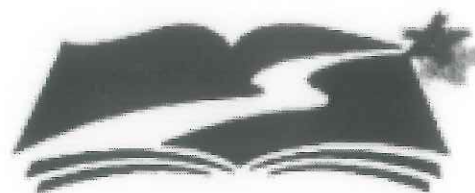
3. 4. Ejemplos de informes de resultados de exámenes criterios de gran escala

A continuación se encuentran algunos ejemplos de formas diversas en que se pueden presentar los informes de resultados, así como las secciones que componen a cada ejemplo.

INFORMES INDIVIDUALES

En la Figura 5 se observa un informe individual de un estudiante evaluado en el área de Artes e Idioma Inglés. El informe está conformado por diversas secciones, entre las cuales se incluye un recuadro en la sección superior izquierda que destaca, a través de una gráfica de cajas y bigotes, el lugar que mantiene el alumno en relación con los estándares de lectura requeridos. En la sección superior derecha, se encuentran dos gráficas de barras que comparan la ejecución del alumno con la de la escuela y el distrito escolar en dos áreas curriculares evaluadas. Finalmente, en la sección inferior se encuentran recomendaciones de actividades que debe realizar el estudiante en cada área curricular evaluada.

Student Report Key



SCHOOL: NEWTOWN ELEM - 000
DISTRICT: NEWTOWN - 00

2005 DELAWARE STUDENT TESTING PROGRAM

English Language Arts Individual Report for

JOHN DOE

Student ID#: 000000

A

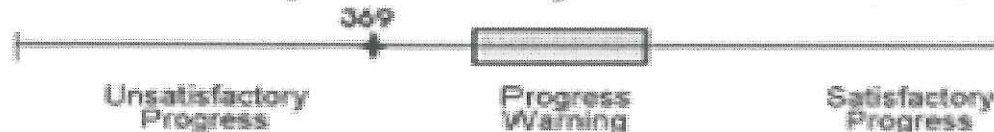
A

GRADE TESTED: 02
TEST DATE: 03/10/05
SAT10 LEVEL/FORM: P2/B
SAT10 NORMS: 2002 PD 14

PROGRESS TOWARDS THE STANDARDS

This test is designed to measure your child's progress in terms of the Delaware Content Standards.

Your Child's Progress in Reading:



B

SCORE COMPARISONS OF GRADE TESTED

Reading

Student 369
School 419
District 419



D

Text-based Writing

3.00
 3.50
 3.50



E

Certain items on the Reading part of the test were administered to a national sample of students. The percentile below represents how your child performed on those items compared to other students in the same grade throughout the country.



INSTRUCTIONAL NEEDS

To improve performance in READING, your child should work on:

- determining meaning by reading more carefully to retell or restate information from the text.
- determining meaning by using context clues, identifying story elements, and organizing important points to understand the text.
- interpreting meaning by drawing conclusions about the central ideas in a text and understanding why a text was written.
- interpreting meaning by using information to make inferences.
- extending meaning by drawing conclusions and using critical thinking to connect and synthesize information within and across text, ideas, and concepts.

To improve performance in WRITING, your child should work on:

- using an effective introduction and closing.
- writing in a consistent style with precise, vivid word choice.
- writing with a clear, logical progression of ideas using smooth transitions.
- including relevant details that are fully elaborated.

F

COPY 01

Please see your child's teacher for more information.

Process No.: 17788999-05ELI22-0000-01110-5
00000

Copyright © 2005 by Delaware Department of Education. Scores based on normative data copyright © 2003 by Harcourt Assessment, Inc. All rights reserved.

Figura 5. Ejemplo de informe de resultados individuales.
Fuente: Delaware Student Testing Program, grade 2. Your Student's Test Results. 2005.

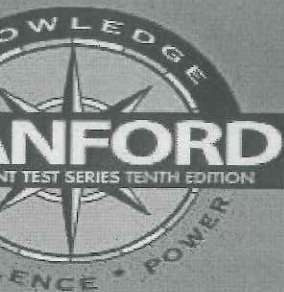
Otro ejemplo de informe individual, es el que se presenta a continuación en la Figura 6. En él se puede observar en la parte superior los datos generales del alumno, así como una nota que indica que el tipo de evaluación realizada fue a nivel nacional. En la parte superior izquierda se encuentra una breve descripción de la ejecución general del alumno; en la parte superior derecha se encuentra un recuadro que indica los puntajes totales en cada subprueba y área evaluadas, con los correspondientes puntajes a nivel nacional y el grado académico al que corresponden los resultados del alumno, de tal manera que se ubique al alumno en los rangos percentilares que le corresponden, según los resultados nacionales, de manera clara y sencilla a través de las gráficas proporcionadas. Finalmente, en la sección inferior del informe, se desglosa cada una de las áreas evaluadas en las subpruebas, indicando los niveles de dominio de áreas curriculares específicas, lo cual será de mucha utilidad para el profesor que imparte cada materia evaluada.

En la Figura 7, se muestra un informe de resultados de la ejecución de un estudiante en comparación con el nivel de dominio de los estándares académicos establecidos, que se supone debe obtener como resultado de la evaluación. Se hace una breve descripción a manera de informe de la ejecución del alumno; también se incluye un recuadro que explica el significado de cada nivel de dominio establecido por los evaluadores. En la esquina superior izquierda se indica de manera gráfica el nivel en el que se encuentra la ejecución del alumno en relación a los estándares de ejecución. En el resto del informe se incluyen recuadros que explican en forma clara y concisa los resultados de la ejecución del alumno en cada una de las áreas del

currículum evaluadas, los cuales son resúmenes narrativos que sirven para dilucidar lo que saben y pueden hacer los estudiantes, y al mismo tiempo sirven como guía para la planeación instruccional. Este tipo de informe se puede hacer a nivel grupal, de distrito o nacional, según se requiera.

En la Figura 8, se muestra un informe de resultados de la ejecución de un estudiante, dirigido a los padres de familia y al estudiante. Al inicio se presenta una breve explicación respecto a la ejecución que tuvo el examinando en comparación con la ejecución de estudiantes del mismo grado escolar que pertenecen al grupo de referencia nacional, y utiliza la gráfica que se encuentra en la sección superior derecha como referencia para dicha explicación. El documento incluye una breve descripción de lo que mide cada sección del examen, según el área curricular de la que se trate; una breve interpretación de los resultados de la ejecución de la alumna, redactada con lenguaje sencillo y conciso; y recomendaciones de actividades específicas que se pueden realizar en casa para ayudar a la alumna a mejorar en las áreas que fueron evaluadas por el examen.

De esta manera se establece una comunicación con los padres de familia, y por ende se despierta en ellos un interés y compromiso por las actividades académicas de sus hijos.



Stanford Achievement Test®,
Eighth Edition

Student's Performance:

Use the *Stanford Achievement* (Stanford 10). This test is a measure of a child's achievement. This report shows a child's performance to the grade across the nation. The report shows ranges within which a child's performance is likely to fall. For example, the 50th Percentile Band spans the range of scores that 50% of students performed as well as or better than nationally in that subject.

This report shows this child's performance in the area tested.

Stanford Achievement Test® Score = 730L

For more information on the use of Lexiles can be found at www.pearson.com.

More information is available to help you match students with reading materials.

	30	30	22
Words	12	12	9
	9	9	5
	9	9	8
ension	18	18	13
	54	54	35
	18	18	12
	18	18	10
	18	18	13
	12	12	11
	20	20	12
	12	12	8
	10	10	4
m Solen	42	42	24
	18	18	20

Student Report | DARCIE L JONES

National Comparison

TEACHER: TILTON
SCHOOL: NEWTOWN ELEMENTARY - 215648
DISTRICT: NEWTOWN

GRADE: 04
TEST DATE: 04/03

AGE: 09 Yrs 08 Mos
STUDENT NO.: 8

Subtests and Totals	Number Possible	Number Correct	Scaled Score	National PR-S	National NCE	Grade Equivalent	AAC Range	National Grade Percentile Bands						
Total Reading	114	82	639	59-5	54.8	5.4	MIDDLE	1	10	30	50	70	90	99
Word Study Skills	30	25	664	76-6	64.9	7.6	HIGH							
Reading Vocabulary	30	22	627	46-5	47.9	4.5	MIDDLE							
Reading Comprehension	54	35	634	53-5	51.6	5.0	MIDDLE							
Total Mathematics	80	56	633	64-6	57.5	5.5	MIDDLE							
Mathematics Problem Solving	48	30	623	54-5	52.1	5.0	MIDDLE							
Mathematics Procedures	32	26	650	74-6	63.5	6.3	HIGH							
Language	48	28	610	39-4	44.1	3.5	MIDDLE							
Language Mechanics	24	15	617	46-5	47.9	4.3	MIDDLE							
Language Expression	24	13	603	36-4	42.5	3.5	MIDDLE							
Spelling	40	30	647	73-6	62.9	6.4	HIGH							
Science	40	30	643	69-6	60.4	6.3	MIDDLE							
Social Science	40	22	607	40-5	44.7	3.5	MIDDLE							
Listening	40	22	608	35-4	41.9	3.4	MIDDLE							
Thinking Skills	190	122	623	56-5	53.2	5.1	MIDDLE							
Basic Battery	322	218	NA	57-5	53.6	5.0	MIDDLE							
Complete Battery	402	270	NA	56-5	53.4	5.0								
								Review Achievement/Ability Comparisons (AACs) when Stanford 10 is administered with the Otis-Lennon School Ability Test, Eighth Edition.						
Otis-Lennon School Ability Test®	Number Possible	Number Correct	SAI	Age PR-S	Scaled Score	Natl Grade PR-S		1	10	30	50	70	90	99
Total	72	33	106	65-6	605	55-5								
Verbal	36	21	112	77-7	613	63-6								
Nonverbal	36	17	102	55-5	597	46-5								

Avg	Above Avg	Clusters	NP	NA	NC	Below Avg	Avg	Above Avg	Clusters	NP	NA	NC	Below Avg	Avg	Above Avg
.	.	Mathematics Procedures	32	32	26	.	.	.	Science (cont.)				.	.	.
.	.	Computation w/Whole Numbers	18	18	14	.	.	.	Form & Function	13	13	9	.	.	.
.	.	Computation with Decimals	8	8	6	.	.	.	Thinking Skills	20	20	16	.	.	.
.	.	Computation with Fractions	6	6	6	.	.	.	Social Science	40	40	22	.	.	.
.	.	Computation in Context	16	16	13	.	.	.	History	10	10	6	.	.	.
.	.	Computation/Symbolic Notation	16	16	13	.	.	.	Geography	10	10	8	.	.	.
.	.	Thinking Skills	16	16	13	.	.	.	Political Science	10	10	6	.	.	.
.	.	Language Mechanics	24	24	15	.	.	.	Economics	10	10	2	.	.	.
.	.	Capitalization	8	8	7	.	.	.	App. of Knowledge/Comp.	14	14	7	.	.	.
.	.	Usage	8	8	3	.	.	.	Org., Summ. & Interp. of Info.	15	15	7	.	.	.
.	.	Punctuation	8	8	5	.	.	.	Determination of Cause/Effect	11	11	8	.	.	.
.	.	Language Expression	24	24	13	.	.	.	Thinking Skills	20	20	11	.	.	.
.	.	Sentence Structure	8	8	4	.	.	.	Listening	40	40	22	.	.	.
.	.	Prewriting	5	5	3	.	.	.	Vocabulary	10	10	3	.	.	.
.	.	Content and Organization	11	11	6	.	.	.	Comprehension	30	30	19	.	.	.
.	.	Thinking Skills	12	12	6	.	.	.	Initial Understanding	8	8	6	.	.	.
.	.	Spelling	40	40	30	.	.	.	Interpretation	12	12	7	.	.	.
.	.	Phonetic Principles	18	18	14	.	.	.	Analysis	7	7	4	.	.	.
.	.	Structural Principles	18	18	7	.	.	.	Organization	2	2	2	.	.	.



STUDENT PERFORMANCE STANDARDS REPORT

DARCIE L JONES

Academic Performance Standards

TEACHER: SMITH - 0000012345
SCHOOL: LAKESIDE ELEMENTARY - 0000000567
DISTRICT: NEWTOWN - 1222222222
GRADE: 04 AGE: 11 Yrs 01 Mos
TEST DATE: 04/03 STUDENT NO.: 1111111111
OTHER INFO.: 0000000008

Standards Connection

Below Basic: Indicates little or no mastery of fundamental knowledge and skills.

Basic: Denotes partial mastery of the knowledge and skills that are fundamental for satisfactory work.

Proficient: Represents solid academic performance, indicating that students are prepared for the next grade.

Advanced: Signifies superior performance beyond grade-level mastery.

About This Student's Performance Report:

Darcie recently took the *Stanford Achievement Test Series, Tenth Edition* (Stanford 10). This test is one measure of this student's achievement. The narratives below describe this student's performance on academic standards that represent the skills and concepts a student should know and be able to do in the given subject areas. The chart to the right shows the Performance Standard category for each subtest reported.

	Reading	Math	Language	Spelling	Science	Soc. Sci.	Listening
Advanced	☐	☐	☐	☐	☐	☐	☐
Proficient	☒	☐	☐	☒	☐	☒	☐
Basic	☐	☐	☒	☐	☒	☐	☒
Below Basic	☐	☒	☐	☐	☐	☐	☐

Teachers get clear information about a student's performance in each content area.

READING

Darcie's Reading score is in the Proficient range. Students whose scores fall in this range are reasonably able to develop an initial understanding of explicitly stated relationships in a variety of reading selections. They can form interpretations of reading passages based on implicit and explicit information in those passages. They perform at a high level with recreational and functional content, but are less successful with textual content.



MATHEMATICS

Darcie's Mathematics score is in the Proficient range. Students whose scores fall in this range have developed a very good understanding of whole numbers and principles of arithmetic. They are beginning to understand the relationship between addition and multiplication and have a good grasp of basic math facts and geometric principles. They understand some of the laws governing chance and have a good understanding of probability.



Narrative summaries provide insight about what students know and can do and guide instructional planning.

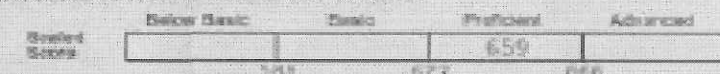
LANGUAGE

Darcie's Language score is in the Proficient range. Students whose scores fall in this range are proficient with the basics of capitalization and punctuation. They can recognize and correct subject-verb agreement, verb tense, and pronoun case. They can combine fragments into sentences, correct phrases with commas and use participles. They are beginning to develop composition skills and reference skills.



SPELLING

Darcie's Spelling score is in the Proficient range. Students whose scores fall in this range are reasonably able to develop an initial understanding of explicitly stated relationships in a variety of reading selections. They can form interpretations of reading passages based on implicit and explicit information in those passages. They perform at a high level with recreational and functional content, but are less successful with textual content.



SCIENCE

Darcie's Science score is in the Proficient range. Students whose scores fall in this range perform well on most of the objectives measured at this level. They do about equally well in earth science, physical science, and life science. Across content areas, they are able to make inferences and draw conclusions based on evidence and analyze data resulting from observations.



SOCIAL SCIENCE

Darcie's Social Science score is in the Proficient range. Students whose scores fall in this range are successful with many of the objectives measured on this test. They have developed broad knowledge in history, geography, civics and government, economics and cultural understanding. They show ability to make inferences and draw conclusions based on various sources of information and have developed some ability to analyze social processes.



LISTENING

Darcie's Listening score is in the Basic range. Students whose scores fall in this range have difficulty understanding explicit and implicit information in listening selections.



STANFORD LEVELFORM: Intermediate to
2002 NORMS: Spring National

Scores based on normative data copyright © 2003 by Harcourt Assessment, Inc. All rights reserved.

This report is also available for a class or district.

000HAPS-0000-12060-3

Figura 7. Ejemplo de informe de resultados individuales comparados con niveles de dominio y estándares de ejecución ideales

Fuente: Stanford Achievement Test series, Tenth Edition. Reporte elaborado en 2003 por Harcourt Assessment, Inc.



Home Report | DARCIE L JONES

Clear, easily understood information, including suggestions for home activities, helps to communicate with parents.

National Comparison

The Learning Snapshot provides a quick look at a student's strengths and needs.

Learning Snapshot

- You have a solid foundation in Spelling, Science and Total Math. Keep working on expanding your skills in these areas.

A Lexile measure is available to provide an easy way to match students to appropriate reading materials.

Lexile Measure = 730L

Information on the use of lexiles can be found at www.lexile.com. Lexiles used with permission.

TEACHER: TILTON
SCHOOL: NEWTOWN ELEMENTARY - 215648

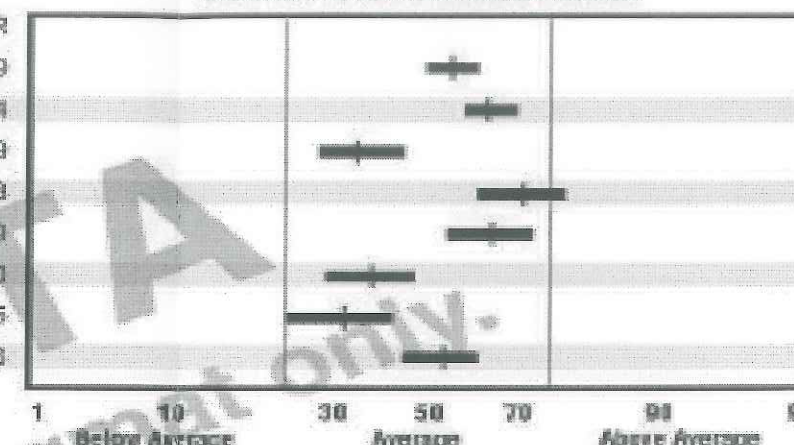
AGE: 09 Yrs 08 Mos

About This Student's Performance Report:

Darcie recently took the *Stanford Achievement Test, Tenth Edition (Stanford 10)*. The graph on the right presents National Grade Percentile Bands, which show your student's performance compared to students in the same grade in a national reference group. Percentile Bands show ranges within which your student's true scores likely fall. For example, a student whose Percentile Band spans the 70th percentile performed as well as or better than 70% of students at subject. The boxes below describe what measures and your student's performance in each. Also included are some suggested activities to engage you and your student as you work together toward continued academic development. This test is only one means for addressing your student's achievement. The school should be contacted for more information about your student's overall performance.

	National PR
Total Reading	59
Total Math	64
Language	39
Spelling	73
Science	60
Social Science	40
Listening	35
Complete Battery	58

National Grade Percentile Bands



TOTAL READING

The Reading subtests measure reading skills such as relating word sounds and spellings, determining word meanings and synonyms, as well as the understanding, interpretation, and analysis of literary, informational, and functional reading selections. Darcie's score is in the Average range for the grade. Print the names of common household objects on small pieces of paper and attach them to the objects. Label each thing as bed, chair, table, floor, etc. Read them with your child often. Leave the labels up for several weeks so your child will see them every day.

SPELLING

The Spelling subtest measures the student's ability to recognize the correct spelling of words, as well as the application of phonetic and structural principles to the identification of correctly spelled words. Darcie's score is in the Average range for the grade. Help your student create a weekly comic strip using as many of his or her spelling words as possible in the characters' conversations. Display your student's work.

LISTENING

The Listening subtest measures student recognition of spoken words and the student's ability to construct meaning from dictated material. Also measured is the student's ability to understand, interpret, and analyze information contained in literary, informational, and functional messages. Darcie's score is in the Average range for the grade. Encourage your student to read challenging grade-level material. Listen to radio and television messages together and discuss what you heard. Ask your student to take a stand on a particular message and then to defend his or her stance.

TOTAL MATHEMATICS

The Mathematics subtests measure problem solving skills involving number sense, operations, patterns and algebra, data and probability, geometry, and measurement concepts. Also measured is the student's fluency with arithmetic operations involving whole numbers, decimals, and fractions. Darcie's score is in the Average range for the grade. Encourage estimation skills by having your student use mental math to track total expenditures on shopping trips. Discuss the meaning of information presented in graphs or charts in newspapers and magazines. Take advantage of home improvement projects to explore measurement concepts like length and area.

SCIENCE

The Science subtest measures the student's understanding of life science, Earth science, physical science, and the nature of science. Also measured is the student's ability to analyze evidence and models, recognize patterns, and compare the forms and functions of organisms. Darcie's score is in the Average range for the grade. Take your student to zoos and museums. Encourage your student to describe the things in the natural world that interest him or her. Read children's science books and magazines with your student.

LANGUAGE

The Language subtest measures the student's application of the language principles that form effective writing including capitalization, punctuation, word usage, sentence structure, organization, composing, and editing. Darcie's score is in the Average range for the grade. Help your student find a pen pal living in a distant place and encourage frequent letter writing. The pen pal could be a relative in another city or state. Talk about family experiences, community events, and school activities that could be related in the letters and messages.

SOCIAL SCIENCE

The Social Science subtest measures the student's achievement in the areas of history, geography, political science, and economics. Also assessed is the student's ability to apply that knowledge and analyze new information. Darcie's score is in the Average range for the grade. Ask your student to draw simple maps. Share and discuss daily news headlines. Encourage your student to read and discuss content in informational, biographical, or historical fiction books.

STANFORD LEVEL FORM: INTERMEDIATE 1/4
2000 NORMS: Spring National

Scores based on normative data copyright © 2003 by Harcourt Assessment, Inc.

COPYRIGHT
FRONT

¡Sí!

For Parents:
This report is also
available in Spanish.

Figura 8. Ejemplo de informe de resultados individuales dirigido a padres de familia

Fuente: Stanford Achievement Test series. Tenth Edition. Reporte elaborado en 2003 por Harcourt Assessment, Inc.

INFORMES GRUPALES

A continuación se explica e ilustra un ejemplo de una de las diversas formas en que se puede presentar un informe de resultados grupal. En la figura 9, se encuentran los resultados de la ejecución de un grupo de estudiantes, en donde se puede observar que hay cuatro secciones principales, diseñadas para responder a las siguientes preguntas:

1. **¿Cuál es el puntaje promedio del grupo en el área curricular evaluada?** En esta sección se encuentra una breve explicación del significado de la puntuación global de la ejecución grupal, así como una gráfica que apoya dicha explicación.
2. **¿Están los estudiantes evaluados cumpliendo con los estándares estatales?** Aquí se dividió a los estudiantes en cuatro grupos para realizar la interpretación de los resultados: los que obtuvieron puntuaciones que los ubicaron en un nivel básico, los que obtuvieron resultados en un nivel eficiente, los que fueron ubicados en un nivel de ejecución acelerada, y los de ejecución de tipo avanzada. Se explica en forma concisa lo que cada nivel significa y el rango de puntuaciones que se utilizó para clasificar dichos niveles, y se interpretan las puntuaciones grupales.
3. **¿Cuáles son los siguientes pasos y qué recomendaciones se sugieren?** En esta sección se brindan sugerencias para mejorar el desempeño y apoyar el desarrollo académico del grupo.
4. **¿Qué ejecución tuvo el grupo evaluado en relación a cada estándar de contenido?** En este ejemplo se explican en forma abreviada los resultados

grupales en cada una de las áreas curriculares que evaluó la prueba, destacando lo que el grupo puede hacer (competencias), y no lo que no puede hacer. Se hace mención de las preguntas específicas que fueron utilizadas para evaluar cada área del currículum, así como del porcentaje de respuestas correctas que obtuvo el grupo en cada una de ellas. Con esta información y la que se tiene de la escuela, distrito escolar y estado, se presenta una gráfica de niveles de dominio alcanzados por el grupo, en comparación con los resultados del ciclo escolar anterior, de toda la escuela, del distrito, así como los del estado.



What Is My Class's Average Reading Score?

Reading Achievement Results

CLASS AVERAGE SCORE

Your class's average score is 404. This is lower than the performance of the third-graders in your school, lower than that of third-graders in your district and lower than that of third-graders statewide.

Your Class's Average Score: 404

Limited Basic Proficient Accelerated Advanced

State Average Score: 416

District Average Score: 411

School Average Score: 411

Last Fall Average Score: 392

The scale scores in this report are used to keep the performance standards (e.g., the proficient level) comparable from year to year even though a new version of the test is used during each testing cycle. Scores in different subjects cannot be compared.

Are Students Meeting State Standards?

HOW YOUR CLASS PERFORMED

This chart compares the percentage of students from your class who scored in each performance level with the percentage of students who scored in each performance level in your school, your district and the state. (Your class is included in the averages for the other groups.)

Students whose scores fall within the proficient range meet state standards. When compared with students throughout the state, a lower percentage of your students are proficient or above.

Note: This report is not intended for teacher evaluations.

Fall performance score includes 21 of your students; spring performance score includes 21 of your students.

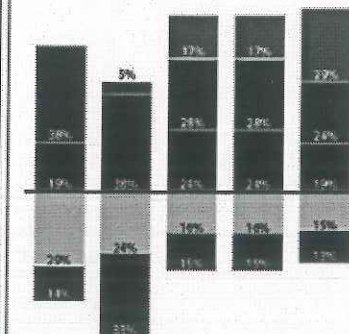
Recommended Next Steps

Your class's performance on Reading does not suggest a relative weakness compared with other classes in your school. However, to support students' growth, you could target Reading Process with activities such as the ones listed below.

- Have students read aloud appropriate-level texts, pausing to ask questions, make predictions and check comprehension.
- Have students use teacher-created reading guides to help them use text features and identify central ideas (guides can also include diagrams for predictions, summaries or comparisons).
- Divide students into mixed-ability groups. Using the jigsaw method, have each group read part of a text. Have the groups report summaries, predictions and conclusions.

% Proficient or Above

37% 43% 69% 69% 71%



Percentages may not sum to totals because of rounding.

WHAT THE PERFORMANCE LEVELS MEAN

Advanced (432-520)

Students go beyond understanding what they read to explain what they like or do not like about a story.

Accelerated (415-431)

Students understand what they read (e.g., fairy tales, folk tales, poetry) and compare stories with other stories.

Proficient (400-414)

Students understand what they read. They try to list important ideas in the texts they read.

Basic (385-399)

Students understand some of what they read. They use strategies (e.g., clues in sentences and paragraphs, knowledge of word parts) to learn new words.

Limited (250-384)

Students may struggle with simple reading tasks (e.g., fairy tales, folk tales, poetry).



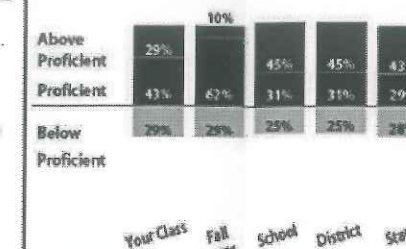
How Did My Class Do on Each Content Standard?

The test has enough questions to report only three levels of performance for each content standard.

NOTE: The released questions in the tables to the right reflect the composition of the test. The bar graphs directly below represent your class's performance on all the questions. Individual test questions give an incomplete picture of student performance, so it is better to focus on the performance data for each standard.

Acquisition of Vocabulary

Students figure out what new words mean by looking at the words and sentences around the new words. They correctly read common words (e.g., *ant*, *leaf*, *pillow*). They use what they know about parts of words (e.g., *can't*, *lovely*, *helpful*) to understand whole words.



Starting in July, you can see some of the actual test questions at www.success.ode.state.oh.us

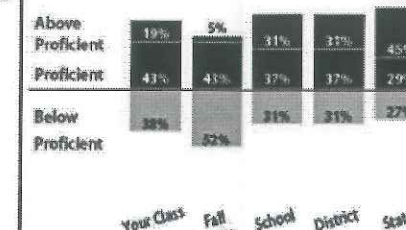
Below you will see your class's performance on the released questions in each content standard.

MULTIPLE-CHOICE QUESTIONS

QUESTION NUMBER	2	3	4	8	26
PERCENT CORRECT	100	95	90	86	81

Reading Process

Students read books by themselves and use reading strategies to help understand what they read. For example, they summarize stories, articles or book chapters by putting information in the right order and by including main ideas and important details.



MULTIPLE-CHOICE QUESTIONS

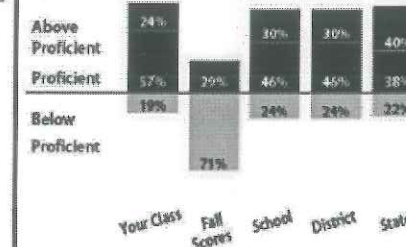
QUESTION NUMBER	10	11	12	21
PERCENT CORRECT	90	95	82	81

CONSTRUCTED-RESPONSE QUESTIONS

QUESTION NUMBER	13	25
AVERAGE POINTS	1.2	1.4
POSSIBLE POINTS	2	4

Informational Text

Students use text features (e.g., titles, index, pictures) to understand informational texts (e.g., maps, instructions, non-fiction). They know how to draw conclusions from information in charts, graphs and diagrams.



MULTIPLE-CHOICE QUESTIONS

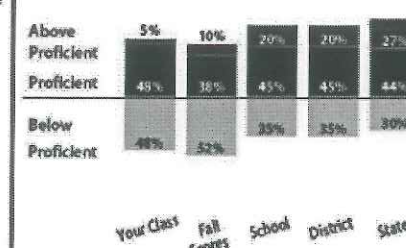
QUESTION NUMBER	7	14	15
PERCENT CORRECT	86	87	76

CONSTRUCTED-RESPONSE QUESTIONS

QUESTION NUMBER	9
AVERAGE POINTS	3.2
POSSIBLE POINTS	4

Literary Text

Students compare and contrast the plots of different stories. They use details from stories to describe characters and setting and can retell what happens. They also explain the features of different types of literature (e.g., fairy tales, folk tales, poems).



MULTIPLE-CHOICE QUESTIONS

QUESTION NUMBER	22	23	24
PERCENT CORRECT	48	62	71

Figura 9. Ejemplo de informe de resultados grupales

Fuente: Departamento de Educación de Ohio (2005). Muestra de "Reporte de resultados grupales" de la Prueba de Aprovechamiento en Lectura, de Ohio.

Conclusiones generales.

La elaboración de los informes de resultados es una etapa del proceso de construcción de exámenes que tiene como requisito un proceso previo de planeación, diseño y aplicación del examen, que arroja productos notables como son: una retícula que hace explícito el universo de contenido cuyo dominio evalúa el test y que define el segmento de dicho universo que es importante evaluar; un documento que justifica dichas decisiones; una tabla de especificaciones del examen que detalla la estrategia evaluativa que se empleará para explorar el dominio que tiene el examinando de ese contenido curricular importante; un prototipo general que especifica las características de las tareas evaluativas mediante las cuales se constatará dicho dominio; una serie de especificaciones de ítems que mantienen una alta congruencia con las decisiones de planeación antes descritas y que logran una coherencia tal que permiten prefigurar un conjunto de ítems que constituyen una muestra representativa del universo de contenido a evaluar, y que garantizan la relación ítem representa currículum, respondiendo de manera rigurosa a las normas técnicas implícitas en el proceso diseñado para su elaboración, y con ello la validez de la prueba; un análisis de los ítems y modelos de examen que nos permite contar con evidencias sobre la calidad técnica del examen, de modo que podamos tener confianza en las versiones de la prueba y efectuar interpretaciones válidas a partir de sus resultados; todo ello ya con suficientes datos empíricos tras la aplicación a gran escala de los modelos de examen generados, mismos que permitieron también realizar un análisis de la ejecución de los examinandos por área de contenido

definida por el currículum.

Los conocimientos y habilidades adquiridos por los participantes en el módulo no solo les permite elaborar los análisis e informes de resultados, lo cual es la meta primaria de su participación, sino que los pone en una posición de privilegio y responsabilidad en su institución, pues ellos serán un recurso fundamental para proporcionar información certera, clara, útil y confiable a distintos usuarios.

Glosario de términos sobre la evaluación del aprendizaje

Agregado. La acumulación de datos similares en un conjunto más grande. Por ejemplo, el cómputo de los mismos puntajes en un examen de lectura de varios grupos, de varias escuelas, en una inspección escolar.

Alineamiento. Correspondencia entre métodos instruccionales, evaluaciones y estándares curriculares.

Aprendizaje. Cambio relativamente permanente en la conducta debido a la experiencia.

Calificación analítica del rasgo. Un procedimiento de la calificación en el que se evalúan de acuerdo a ciertos criterios las ejecuciones en rasgos o dimensiones seleccionadas, en las cuales cada rasgo recibe un puntaje separado. Este método se usa a menudo en la evaluación de muestras de escritura y de otras ejecuciones de examinandos.

Calificación holística. Asignación de un puntaje global a una ejecución de acuerdo con ciertos criterios.

Calificación por curva. Una práctica común de asignar las calificaciones en base al supuesto de que las frecuencias de respuesta asociadas con el rasgo que se mide tienen una distribución en forma de campana y que debe haber tantos seises como dieces (ver curva normal)

Confiabilidad. El grado en que los resultados de una evaluación son fidedignos, y estables. La confiabilidad indica la consistencia de los puntajes entre evaluadores, en

el tiempo, o en tareas diferentes que miden la misma cosa o dentro de una sola prueba. Así, puede ser de ocasión a ocasión, de juez a juez, de prueba a prueba, de ítem a ítem, etc.

Contenido. De lo que trata una comunicación sobre algo; es decir, la sustancia, asunto, mensaje. También todo lo incluido en una colección (por ejemplo un programa, un curso, etc.).

Curriculum. Curso integrado de estudios académicos.

Curva Normal. Familia de curvas simétricas, en forma de campana, cuya forma se cree que subyace comúnmente a muchos rasgos medidos por las evaluaciones psicológicas (ver Calificación por curva)

Desagregación. Descomposición de datos sumarios en partes constituyentes. Por ejemplo, reportar los puntajes obtenidos en ciencia por una escuela, por género o nivel socioeconómico.

Distractores. Opciones incorrectas que se ofrecen en un reactivo al examinando con el objeto de poner a prueba su dominio de un contenido.

Especificaciones de reactivos. Descripción técnica de las características relevantes, desde el punto de vista psicométrico y disciplinario, que debe tener un reactivo que será elaborado para una prueba.

Estándar de ejecución. La descripción de conocimientos y/o habilidades esperados en diferentes niveles de desarrollo a fin de que un estudiante demuestre su dominio del contenido.

Estandarización. Un conjunto de procedimientos consistentes para construir, administrar y calificar una prueba. La meta de la estandarización es asegurar que todos los estudiantes son evaluados bajo condiciones uniformes de modo que la interpretación de su ejecución sea comparable y no influida por condiciones diferentes no pretendidas.

Evaluación. Estimación del valor de algo. Acto de juzgar las evidencias para fundamentar decisiones. Juicio con respecto a la calidad, valor, o mérito de una respuesta, política, programa, producto, o ejecución, basado en criterios establecidos. Las evaluaciones normalmente incluyen fuentes múltiples de información. Los términos "Medición" y "evaluación" son a menudo usados incorrectamente de manera intercambiable.

Evaluación de alto impacto. Cualquier programa de evaluación cuyos resultados tienen consecuencias importantes para los estudiantes, maestros, y/o directivos escolares. Tales impactos pueden incluir promoción, certificación, graduación, aceptación o rechazo del programa, acceso a recursos, etc.

Evaluación de ejecución. La observación y calificación directa de la ejecución o trabajo del estudiante, mediante el juicio profesional sobre la calidad de esa ejecución. Una evaluación de la ejecución de calidad posee criterios preestablecidos de ejecución.

Evaluación de gran escala. Evaluaciones para los propósitos y audiencias fuera del aula. Normalmente se estandarizan las evaluaciones a gran escala en alguna medida y se aplican al mismo tiempo a grandes números de estudiantes. Los

propósitos para la evaluación de gran escala incluyen asumir responsabilidad, evaluación de programas, certificación de competencias y selección de estudiantes para programas especiales. La evaluación de gran escala puede guiar las decisiones instruccionales en el aula, pero normalmente ése no es su propósito primario.

Evaluación referida a un criterio. Comparar la actuación de un estudiante con una descripción de la ejecución deseada o criterio. Por ejemplo, se esperaría que un estudiante tenga correctos tres de los cuatro ítems que miden un rasgo particular. En el contexto educativo el criterio usual es el currículum. Este método se contrasta a menudo con la evaluación referida a una norma.

Evaluación referida a una norma. Comparar la ejecución de un estudiante o el resultado de una prueba, con aquéllos de otros estudiantes similares. Por ejemplo: un estudiante teclea mejor que el 80% de los otros estudiantes examinados. Sin embargo, ello no nos dice qué tan rápido escribe realmente el estudiante. Este tipo de evaluación frecuentemente se contrasta con la evaluación referida a un criterio.

Examen criterial. Instrumento para evaluar el aprendizaje cuyos resultados permiten comparar la ejecución de un examinando contra un criterio de ejecución preestablecido, usualmente el nivel de dominio del currículum. Sus resultados nos dicen qué es capaz o no de hacer un examinando en un campo del conocimiento determinado.

Guía para calificar. Un conjunto de criterios establecidos para calificar la ejecución del estudiante en un ítem específico (ver rúbrica).

Ítem o reactivo. Tarea evaluativa. Componente de un examen que explora el dominio que tiene un examinando de un contenido específico.

Medición. El proceso de cuantificar, describir o recoger información sobre la ejecución, de conformidad con ciertas reglas, criterios o teorías.

Nivel de ejecución. Estándar de logro establecido, calidad de ejecución o grado de dominio.

Objetivo. Que no está influido por la emoción, conjetura, o prejuicio; es decir, basado en fenómenos observables, que se presentan con relativa independencia del sujeto o por acuerdo intersubjetivo. Por ejemplo, véase rúbrica (Ver también subjetivo).

Psicometría. Rama de la psicología que estudia las mediciones psicológicas (de procesos cognitivos, emociones, etc.).

Puntaje crudo. Generalmente, el número de ítems correctos (o total de puntos logrados) en un examen.

Puntaje escalar. Derivado de los puntajes crudos (el número total de puntos logrados) para igualar varios modelos o versiones de una prueba.

Puntos de corte. Puntos de división a lo largo de una escala utilizada para establecer niveles de ejecución.

Rango percentil. El porcentaje de estudiantes con calificaciones por debajo de un puntaje crudo o escalar. Por ejemplo, si el 50% de los examinados en una prueba dada están debajo de un cierto punto, ese punto tiene un rango percentil de 50.

Reactivo muestra. Componente de una especificación de reactivos que ilustra la aplicación de los elementos que la integran.

Respuesta construida. Un tipo de ítem de una prueba que requiere una demostración más completa de habilidades y conocimientos del estudiante de la que se requiere en modos de respuesta tradicionales como la opción múltiple, falso/verdadero o relación de columnas.

Retícula. Modelo gráfico que representa un conjunto de contenidos y las relaciones de servicios entre ellos.

Rúbrica. Un conjunto de criterios establecidos para calificar o evaluar la ejecución del estudiante en una tarea evaluativa. Puede utilizarse una sola rúbrica para calificar múltiples tareas similares. Por ejemplo, podría usarse la misma rúbrica para calificar varios experimentos diferentes en ciencia (Ver guía para calificar).

Subjetivo. Particular a un individuo dado; personal. Por ejemplo, las calificaciones en el aula tienden a otorgarse bajo la discreción del maestro (Ver objetivo).

Tabla de especificaciones del examen. Muestra el conjunto de contenidos que representa el universo de medida y sintetiza, usualmente de manera tabular, el proceso de planeación de una prueba. Indica también el número de especificaciones de ítems, así como el número, orden de presentación y tipo de reactivos, entre otros aspectos del instrumento.

Transferencia del aprendizaje. La capacidad para aplicar habilidades, procesos, o contenidos, previamente aprendidos, a situaciones nuevas o diferentes.

Validez. Grado en que una prueba mide lo que dice. Por ejemplo, una evaluación válida de la solución de un problema matemático, mediría la habilidad de los estudiantes para resolver ese problema, no la habilidad para leer y comprender el

problema. Técnicamente, se refiere a la obtención de evidencias que soportan las inferencias basadas en los puntajes obtenidos en el test. Las principales evidencias que se buscan son las relacionadas con el contenido de la prueba, con un criterio que concurre o se predice, o con el constructo que mide el instrumento. Es el indicador más importante de la calidad de un examen.

Bibliografía

- Abad, F. J. ; Atencia, J.; García, C.; Hontangas, P.; Olea, J.; Ponsoda, V.; Revuelta, J.; Suero, M.; y Ximénez, C. (s/f). Establecimiento de los puntos de corte. Proyecto de innovación docente: ayuda a la creación de exámenes. Departamento de psicología social y metodología. Facultad de Psicología. Universidad Autónoma de Madrid. Consultado el 14 de agosto de 2007 en: <http://www.uam.es/docencia/ace/26-corte.htm>
- Berk, R. (1986). A Consumer's Guide to Setting Performance Standards on Criterion-Referenced Tests. Review of Educational Research. Vol. 56, no. 1, pp. 137-172.
- Condelli, L. (2005). Demonstrating Results: developing state and local report cards for adult education. American institutes for research. Washington, DC. Guía preparada para el proyecto: Enhancing performance through accountability. Contract # ED-04-CO-0025/0001. U.S. Department of Education, Office of Vocational and Adult Education, Division of Adult Education and Literacy.
- Contreras L. A. (2000). Desarrollo y Pilotaje de un Examen de Español para la Educación Primaria en Baja California. Trabajo de Tesis presentado para la obtención del grado de Maestro en Ciencias Educativas en el Instituto de

Investigación y Desarrollo Educativo de la Universidad Autónoma de Baja California pg. 15.

Contreras. L. A. y Backhoff, E. (2004). Metodología para elaborar exámenes criterios alineados con el currículo. En: Castañeda, S. Educación, aprendizaje y cognición. Teoría en la práctica. México: Manual Moderno. Pags. 155 - 174

Gómez A., B. (1990). Evaluación criterial: una metodología útil para diagnosticar el nivel de aprendizaje de los alumnos. Madrid, España: Narcea, D. L. 55 pp.

Hambleton, R. K. y Eignor, D. R. (1980). Competency test development, validation and standard setting. En R. M. Jaeger & C. K. Tittle (eds.) Minimum achievement testing: motives, models, measures and consequences (pp. 367-396). Berkley, Ca.: Mc Cutchan.

Hambleton, R. K. (1988). Criterion-referenced measurement. Educational research, methodology, and measurement. An international Handbook. Keeves, J. P. Edit. Australia: 1988. pp. 277-281.

Jaeger, R. M., (1989). Certification of student competence. Educational measurement. Edit. Linn, R. Estados Unidos: American council on Education and the Oxy Press. pp. 485-514.

Jaeger, R. M. y Tucker, C. G. (1998). Analyzing, disaggregating, reporting, and interpreting student's achievement test results: a guide to practice for Title I and beyond. Estados Unidos: Council of Chief State School Officers

Livingston, S. y Zieky, M. (2004). Extractos de Puntuaciones para Aprobar. Manual para establecer normas de rendimiento referentes a exámenes académicos y ocupacionales. Estados Unidos: Educational Testing Service pp. 3-13.

Márquez, M. C. E. (2005). Aplicación y validación de in método para determinar el punto de corte del Examen de Egreso de Inglés de la Universidad Autónoma de Baja California. Trabajo de Tesis presentado para la obtención del grado de Maestra en Ciencias Educativas en el Instituto de Investigación y Desarrollo Educativo de la Universidad Autónoma de Baja California.

Martínez R. F. (2007). El aprendizaje en tercero de primaria en México: español, matemáticas, ciencias naturales y ciencias sociales. Resumen ejecutivo basado en: Backhoff, E. Andrade, E., Sánchez, A. y Peón, M. (2007). El Aprendizaje en Tercero de Primaria en México: español, Matemáticas, Ciencias Naturales y Ciencias Sociales. INEE: México pp. 29-30.

Nitko, A.J. (1994). *A Model for Developing Curriculum-Driven Criterion-Referenced and Norm-Referenced National Examinations for Certification and Selection of Students*. Ponencia presentada en la Conferencia Internacional sobre Evaluación y Medición Educativas, de la Asociación para el Estudio de la Evaluación Educativa en Sudáfrica (ASSESA).

Popham, J. (1990). *Modern Educational Measurement. A Practitioner's Perspective*. USA. MA. Allyn and Bacon.

Roeber, E. D. (1995). *Critical Issue: reporting assessment results*. North Central Regional Educational Laboratory. Consultado el 06 de septiembre de 2005 en <http://www.ncrel.org/sdrs/areas/issues/methods/assment/as600.htm>

Anexo 2

Formato para elaborar informes de resultados.

Anexo 2. FORMATO PARA ELABORAR INFORMES DE RESULTADOS

Lugar:	Fecha:
INFORMACIÓN GENERAL	
Nombre de la Institución educativa evaluada:	
Descripción de la Institución:	
Metas que persigue el reporte:	
Descripción del contexto y propósito de la evaluación:	
Descripción de la población evaluada:	
Descripción de los destinatarios de la información:	
Objetivos del informe de resultados:	
Materia o área curricular evaluada:	

ANÁLISIS DE LA EJECUCIÓN DE LOS EXAMINADOS

Resultados del ítem análisis:

Distribución de porcentajes de aciertos por área curricular:

Tendencias por área curricular respecto al año anterior:

Análisis de la relación entre la ejecución de los examinados y variables de contexto:

INFORMACIÓN GRÁFICA

Gráficas, figuras y tablas:

CONCLUSIONES

Limitaciones encontradas:

Recomendaciones a partir de los resultados (Usos de la información):

Conclusiones generales y sugerencias:

Anexo 3

**Ejemplos para ilustrar la aplicación de los procedimientos y
criterios para el análisis de datos y la elaboración de informes
de resultados.**

Anexo 3. EJEMPLOS PARA ILUSTRAR LA APLICACIÓN DE LOS PROCEDIMIENTOS Y CRITERIOS PARA EL ANÁLISIS DE DATOS Y LA ELABORACIÓN DE INFORMES DE RESULTADOS.

A continuación se encuentran algunos ejemplos de formas diversas en que se pueden presentar los informes de resultados, así como las secciones que componen a cada ejemplo.

INFORMES INDIVIDUALES

En la Figura 1 se observa un informe individual de un estudiante evaluado en el área de Artes e Idioma Inglés. El informe está conformado por diversas secciones, entre las cuales se incluye un recuadro en la sección superior izquierda que destaca, a través de una gráfica de cajas y bigotes, el lugar que mantiene el alumno en relación con los estándares de lectura requeridos. En la sección superior derecha, se encuentran dos gráficas de barras que comparan la ejecución del alumno con la de la escuela y el distrito escolar en 2 áreas curriculares evaluadas. Finalmente, en la sección inferior se encuentran recomendaciones de actividades que debe realizar el estudiante en cada área curricular evaluada.

Student Report Key



2005 DELAWARE STUDENT TESTING PROGRAM

English Language Arts Individual Report for

JOHN DOE

Student ID#: 000000

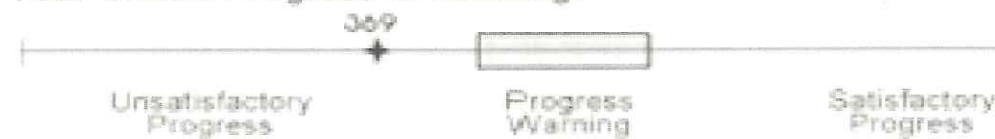
SCHOOL: NEWTOWN ELEM - 000

DISTRICT: NEWTOWN - 00

PROGRESS TOWARDS THE STANDARDS

This test is designed to measure your child's progress in terms of the Delaware Content Standards.

Your Child's Progress in Reading:

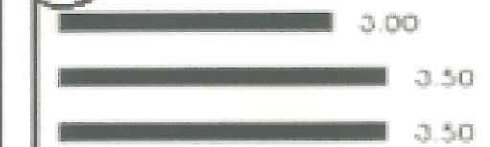


SCORE COMPARISONS OF GRADE TESTED

Reading



Text-based Writing



E

Certain items on the Reading part of the test were administered to a national sample of students. The percentile below represents how your child performed on those items compared to other students in the same grade throughout the country.



INSTRUCTIONAL NEEDS

To improve performance in READING, your child should work on:

- determining meaning by reading more carefully to retell or restate information from the text
- determining meaning by using context clues, identifying story elements, and organizing important points to understand the text
- interpreting meaning by drawing conclusions about the central ideas in a text and understanding why a text may be written
- interpreting meaning by using information to make inferences
- extending meaning by drawing conclusions and using critical thinking to connect and synthesize information within and across text, ideas, and concepts

To improve performance in WRITING, your child should work on:

- using an effective introduction and closing
- writing in a consistent style with precise, vivid word choice
- writing with a clear, logical progression of ideas using smooth transitions
- including relevant details that are fully elaborated

COPY 01

Please see your child's teacher for more information.

Process No: 177005224-5E1 122-0000-011 10-5
00000

Copyright © 2007 by Delaware Department of Education. Scores based on normative data copyright © 2005 by Harcourt Assessment, Inc. All rights reserved.

Figura 1. Ejemplo de informe de resultados individuales.

Fuente: Delaware Student Testing Program, grade 2. Your Student's Test Results. 2005.

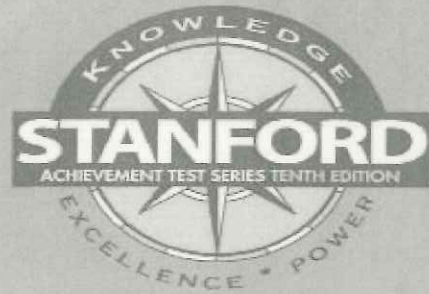
Otro ejemplo de informe individual, es el que se presenta a continuación en la Figura 2. En él se puede observar en la parte superior los datos generales del alumno, así como una nota que indica que el tipo de evaluación realizada fue a nivel nacional. En la parte superior izquierda se encuentra una breve descripción de la ejecución general del alumno; en la parte superior derecha se encuentra un recuadro que indica los puntajes totales en cada subprueba y área evaluadas, con los correspondientes puntajes a nivel nacional y el grado académico al que corresponden los resultados del alumno, de tal manera que se ubique al alumno en los rangos percentilares que le corresponden, según los resultados nacionales, de manera clara y sencilla a través de las gráficas proporcionadas. Finalmente, en la sección inferior del informe, se desglosa cada una de las áreas evaluadas en las subpruebas, indicando los niveles de dominio de áreas curriculares específicas, lo cual será de mucha utilidad para el profesor que imparte cada materia evaluada.

En la Figura 3, se muestra un informe de resultados de la ejecución de un estudiante en comparación con el nivel de dominio de los estándares académicos establecidos, que se supone debe obtener como resultado de la evaluación. Se hace una breve descripción a manera de informe de la ejecución del alumno; también se incluye un recuadro que explica el significado de cada nivel de dominio establecido por los evaluadores. En la esquina superior izquierda se indica de manera gráfica el nivel en el que se encuentra la ejecución del alumno en relación a los estándares de ejecución. En el resto del informe se incluyen recuadros que explican en forma clara y concisa los resultados de la ejecución del alumno en cada una de las áreas del currículum evaluadas, los cuales son resúmenes narrativos que sirven para dilucidar

lo que saben y pueden hacer los estudiantes, y al mismo tiempo sirven como guía para la planeación instruccional. Este tipo de informe se puede hacer a nivel grupal, de distrito o nacional, según se requiera.

En la Figura 4, se muestra un informe de resultados de la ejecución de un estudiante, dirigido a los padres de familia y al estudiante. Al inicio se presenta una breve explicación respecto a la ejecución que tuvo el examinando en comparación con la ejecución de estudiantes del mismo grado escolar que pertenecen al grupo de referencia nacional, y utiliza la gráfica que se encuentra en la sección superior derecha como referencia para dicha explicación. El documento incluye una breve descripción de lo que mide cada sección del examen, según el área curricular de la que se trate; una breve interpretación de los resultados de la ejecución de la alumna, redactada con lenguaje sencillo y conciso; y recomendaciones de actividades específicas que se pueden realizar en casa para ayudar a la alumna a mejorar en las áreas que fueron evaluadas por el examen.

De esta manera se establece una comunicación con los padres de familia, y por ende se despierta en ellos un interés y compromiso por las actividades académicas de sus hijos.



with Otis-Lennon School Ability Test®,
Eighth Edition

About This Student's Performance:

Darcie recently took the *Stanford Achievement Test, Tenth Edition* (Stanford 10). This test is one measure of this child's achievement. This report compares this child's performance to students in the same grade across the nation. Percentile Bands show ranges within which this child's true scores likely fall. For example, a student whose Percentile Band spans the 70th percentile performed as well as or better than 70% of students nationally in that subject - but not as well as 30% of students.

The chart below shows this child's performance in each subject area tested.

Lexile measure = 730L

Information on the use of Lexiles can be found at www.lexile.com.

Lexile is used with permission.

A Lexile measure is available to provide an easy way to match students to appropriate reading materials.

Reading Vocabulary	30	30	22
Synonyms	12	12	9
Multiple Meaning Words	9	9	5
Context Clues	9	9	8
Thinking Skills	18	18	13
Reading Comprehension	54	54	35
Literary	18	18	12
Informational	18	18	10
Functional	18	18	13
Initial Understanding	12	12	11
Interpretation	20	20	12
Critical Analysis	12	12	8
Strategies	10	10	4
Thinking Skills	42	42	24
Mathematics Problem Solving	48	48	30
Number Sense & Operations	24	24	16
Patterns/Relationships/Algebra	6	6	6
Data, Statistics & Probability	8	8	4
Geometry & Measurement	10	10	4
Communication & Representation	6	6	2
Estimation	8	8	5
Mathematical Connections	21	21	13
Reasoning & Problem Solving	13	13	10
Thinking Skills	40	40	26

STANFORD LEVEL/FORM: INTERMEDIATE 1/A
2002 NORMS: Spring National

Student Report | DARCIE L JONES

TEACHER: TILTON
SCHOOL: NEWTOWN ELEMENTARY - 215648
DISTRICT: NEWTOWN

GRADE: 04
TEST DATE: 04/03

AGE: 09 Yrs 08 Mos
STUDENT NO.: 8

Subtests and Totals	Number Possible	Number Correct	Scaled Score	National PR-S	National NCE	Grade Equivalent	AAC Range	National Grade Percentile Bands						
								1	10	30	50	70	90	99
Total Reading	114	82	639	59-5	54.8	5.4	MIDDLE							
Word Study Skills	30	25	664	76-6	64.9	7.6	HIGH							
Reading Vocabulary	30	22	627	46-5	47.9	4.5	MIDDLE							
Reading Comprehension	54	35	634	53-5	51.6	5.0	MIDDLE							
Total Mathematics	80	56	633	64-6	57.5	5.5	MIDDLE							
Mathematics Problem Solving	48	30	623	54-5	52.1	5.0	MIDDLE							
Mathematics Procedures	32	26	650	74-6	63.5	6.3	HIGH							
Language	48	28	610	39-4	44.1	3.5	MIDDLE							
Language Mechanics	24	15	617	46-5	47.9	4.3	MIDDLE							
Language Expression	24	13	603	36-4	42.5	3.5	MIDDLE							
Spelling	40	30	647	73-6	62.9	6.4	HIGH							
Science	40	30	643	69-6	60.4	6.3	MIDDLE							
Social Science	40	22	607	40-5	44.7	3.5	MIDDLE							
Listening	40	22	608	35-4	41.9	3.4	MIDDLE							
Thinking Skills	190	122	623	56-5	53.2	5.1	MIDDLE							
Basic Battery	322	218	NA	57-5	53.6	5.0	MIDDLE							
Complete Battery	402	270	NA	56-5	53.4	5.0	MIDDLE							
								Review Achievement/Ability Comparisons (AACs) when Stanford 10 is administered with the Otis-Lennon School Ability Test, Eighth Edition.						
Otis-Lennon School Ability Test®	Number Possible	Number Correct	SAI	Age PR-S	Scaled Score	Natl Grade PR-S								
Total	72	38	106	65-6	605	55-5								
Verbal	36	21	112	77-7	613	63-6								
Nonverbal	36	17	102	55-5	597	46-5								

Teachers get specific information about skills tested.

Clusters	NP	NA	NC	Below Avg	Avg	Above Avg	Clusters	NP	NA	NC	Below Avg	Avg	Above Avg
Mathematics Procedures	32	32	26				Science (cont.)						
Computation w/Whole Numbers	18	18	14				Form & Function	13	13	9			
Computation with Decimals	8	8	6				Thinking Skills	20	20	16			
Computation with Fractions	6	6	6				Social Science	40	40	22			
Computation in Context	16	16	13				History	10	10	6			
Computation/Symbolic Notation	16	16	13				Geography	10	10	8			
Thinking Skills	16	16	13				Political Science	10	10	6			
Language Mechanics	24	24	15				Economics	10	10	2			
Capitalization	8	8	7				App. of Knowledge/Comp.	14	14	7			
Usage	8	8	3				Org., Summ. & Interp. of Info.	15	15	7			
Punctuation	8	8	5				Determination of Cause/Effect	11	11	8			
Language Expression	24	24	13				Thinking Skills	20	20	11			
Sentence Structure	8	8	4				Listening	40	40	22			
Prewriting	5	5	3				Vocabulary	10	10	3			
Content and Organization	11	11	6				Comprehension	30	30	19			
Thinking Skills	12	12	6				Initial Understanding	8	8	6			
Spelling	40	40	30				Interpretation	12	12	7			
Phonetic Principles	18	18	14				Analysis	7	7	4			
Structural Principles	10	10	7				Strategies	3	3	2			
No Mistake	7	7	7										
Homophones	5	5	2										
Science	40	40	30										
Life	11	11	9										
Physical	11	11	6										
Earth	11	11	10										
Nature of Science	7	7	5										
Models	14	14	11										
Constancy	13	13	10										

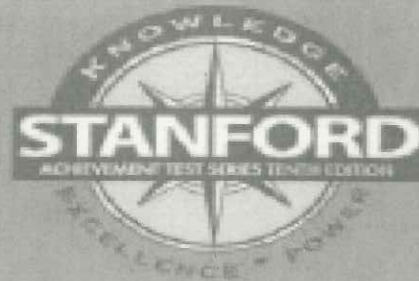
C = Content Cluster P = Process Cluster

Scores based on normative data copyright © 2003 by Harcourt Assessment, Inc.

COPY 01
PROCESS NO. 103000255-00CSRCC-M109-11100-3

Figura 2. Ejemplo de informe de resultados individuales comparados con resultados nacionales

Fuente: Stanford Achievement Test series. Tenth Edition. Reporte elaborado en 2003 por Harcourt Assessment, Inc.



STUDENT PERFORMANCE STANDARDS REPORT

DARCIE L JONES

Academic Performance Standards

TEACHER: SMITH - 0000012345
SCHOOL: LAKESIDE ELEMENTARY - 0000000567
DISTRICT: NEWTOWN - 1222222222
GRADE: 04 AGE: 11 Yrs 01 Mos
TEST DATE: 04/03 STUDENT NO.: 1111111111
OTHER INFO.: 0000000000

Standards Connection

- Below Basic:** Indicates little or no mastery of fundamental knowledge and skills.
- Basic:** Denotes partial mastery of the knowledge and skills that are fundamental for satisfactory work.
- Proficient:** Represents solid academic performance, indicating that students are prepared for the next grade.
- Advanced:** Signifies superior performance beyond grade-level mastery.

About This Student's Performance Report:

Darcie recently took the *Stanford Achievement Test Series, Tenth Edition* (Stanford 10). This test is one measure of this student's achievement. The narratives below describe this student's performance on academic standards that represent the skills and concepts a student should know and be able to do in the given subject areas. The chart to the right shows the Performance Standard category for each subtest reported.

	Performance Standard Category						
	Reading	Math	Language	Spelling	Science	Soc. Sci.	Listening
Advanced	☐	☐	☐	☐	☐	☐	☐
Proficient	☒	☐	☐	☒	☐	☒	☐
Basic	☐	☐	☒	☐	☒	☐	☒
Below Basic	☐	☒	☐	☐	☐	☐	☐

Teachers get clear information about a student's performance in each content area.

READING

Darcie's Reading score is in the Proficient range. Students whose scores fall in this range are reasonably able to develop an initial understanding of explicitly stated relationships in a variety of reading selections. They can form interpretations of reading passages based on implicit and explicit information in those passages. They perform at a high level with recreational and functional content, but are less successful with textual content.



MATHEMATICS

Darcie's Mathematics score is in the Proficient range. Students whose scores fall in this range have developed a very good understanding of whole numbers and principles of arithmetic. They are beginning to understand the relationship between addition and multiplication and have a good grasp of basic math facts and geometric principles. They understand some of the laws governing chance and have a good understand

Scaled Score
Below Basic

Narrative summaries provide insight about what students know and can do and guide instructional planning.

LANGUAGE

Darcie's Language score is in the Proficient range. Students whose scores fall in this range are proficient with the basics of capitalization and punctuation. They can recognize and correct subject-verb agreement, verb tense, and pronoun case. They can combine fragments into sentences, connect phrases with commas and use participles. They are beginning to develop composition



SPELLING

Darcie's Spelling score is in the Proficient range. Students whose scores fall in this range are reasonably able to develop an initial understanding of explicitly stated relationships in a variety of reading selections. They can form interpretations of reading passages based on implicit and explicit information in those passages. They perform at a high level with recreational and functional content, but are less successful with textual content.



SCIENCE

Darcie's Science score is in the Proficient range. Students whose scores fall in this range perform well on most of the objectives measured at this level. They do about equally well in earth science, physical science, and life science. Across content areas, they are able to make inferences and draw conclusions based on evidence and analyze data resulting from observations.



SOCIAL SCIENCE

Darcie's Social Science score is in the Proficient range. Students whose scores fall in this range are successful with many of the objectives measured on this test. They have developed broad knowledge in history, geography, civics and government, economics and cultural understanding. They show ability to make inferences and draw conclusions based on various sources of information and have developed some ability to analyze social processes.



LISTENING

Darcie's Listening score is in the Basic range. Students whose scores fall in this range have difficulty understanding explicit and implicit information in listening selections.



STANFORD LEVELFORM: Intermediate I/A
2002 NORMS: Spring National

Scores based on normative data copyright © 2003 by Harcourt Assessment, Inc. All rights reserved.

This report is also available for a class or district.

000HAPS-0000-1 2050-3

Figura 3. Ejemplo de informe de resultados individuales comparados con niveles de dominio y estándares de ejecución ideales



Home Report | DARCIE L JONES

Clear, easily understood information,
including suggestions for home activities,
helps to communicate with parents.

National Comparison

*The Learning Snapshot
provides a quick look at a
student's strengths and needs.*

Learning Snapshot

- You have a solid foundation in Spelling, Science and Total Math. Keep working on expanding your skills in these areas.

*A Lexile measure is available to
provide an easy way to match students
to appropriate reading materials.*

Lexile Measure = 730L

Information on the use of lexiles can be found at www.lexia.com. Lexiles used with permission.

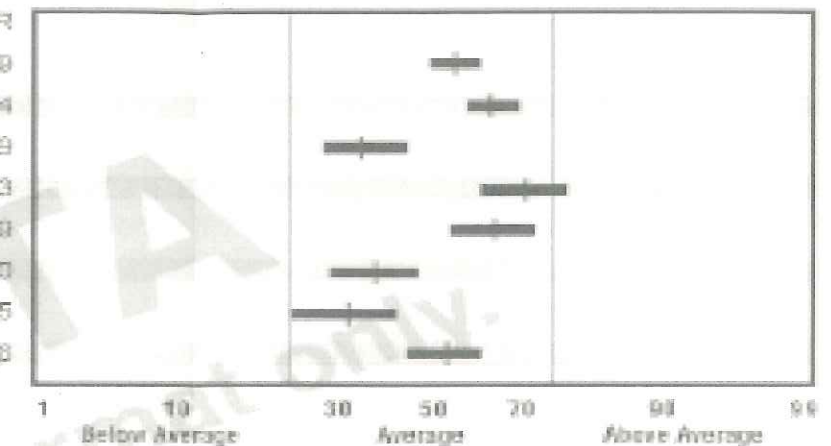
About This Student's Performance Report:

Darcie recently took the *Stanford Achievement Test, Tenth Edition (Stanford 10)*. The graph on the right presents National Grade Percentile Bands, which show your student's performance compared to students in the same grade in a national reference group. Percentile Bands show ranges within which your student's true scores likely fall. For example, a student whose Percentile Band spans the 70th percentile performed as well as or better than 70% of students at subject. The boxes below describe what measures and your student's performance in each. Also included are some suggested activities to engage you and your student as you work together toward continued academic development. This test is only one means for addressing your student's achievement. The school should be contacted for more information about your student's overall performance.

National PR

Total Reading	59
Total Math	64
Language	39
Spelling	73
Science	60
Social Science	40
Listening	35
Complete Battery	58

National Grade Percentile Bands



TOTAL READING

The Reading subtests measure reading skills such as relating word sounds and spellings, determining word meanings and synonyms, as well as the understanding, interpretation, and analysis of literary, informational, and functional reading selections. Darcie's score is in the Average range for the grade. Print the names of common household objects on small pieces of paper and attach them to the objects. Label each thing as bed, chair, table, door, etc. Read them with your child often. Leave the labels up for several weeks so your child will see them every day.

SPELLING

The Spelling subtest measures the student's ability to recognize the correct spelling of words, as well as the application of phonetic and structural principles to the identification of correctly spelled words. Darcie's score is in the Average range for the grade. Help your student create a weekly comic strip using as many of his or her spelling words as possible in the characters' conversations. Display your student's work.

LISTENING

The Listening subtest measures student recognition of spoken words and the student's ability to construct meaning from dictated material. Also measured is the student's ability to understand, interpret, and analyze information contained in literary, informational, and functional messages. Darcie's score is in the Average range for the grade. Encourage your student to read challenging grade-level material. Listen to radio and television messages together and discuss what you heard. Ask your student to take a stand on a particular message and then to defend his or her stance.

TOTAL MATHEMATICS

The Mathematics subtests measure problem solving skills involving number sense, operations, patterns and algebra, data and probability, geometry, and measurement concepts. Also measured is the student's fluency with arithmetic operations involving whole numbers, decimals, and fractions. Darcie's score is in the Average range for the grade. Encourage estimation skills by having your student use mental math to track total expenditures on shopping trips. Discuss the meaning of information presented in graphs or charts in newspapers and magazines. Take advantage of home improvement projects to explore measurement concepts like length and area.

SCIENCE

The Science subtest measures the student's understanding of life science, Earth science, physical science, and the nature of science. Also measured is the student's ability to analyze evidence and models, recognize patterns, and compare the forms and functions of organisms. Darcie's score is in the Average range for the grade. Take your student to zoos and museums. Encourage your student to describe the things in the natural world that interest him or her. Read children's science books and magazines with your student.

LANGUAGE

The Language subtest measures the student's application of the language principles that form effective writing including capitalization, punctuation, word usage, sentence structure, organization, composing, and editing. Darcie's score is in the Average range for the grade. Help your student find a pen pal living in a distant place and encourage frequent letter writing. The pen pal could be a relative in another city or state. Talk about family experiences, community events, and school activities that could be related in the letters and messages.

SOCIAL SCIENCE

The Social Science subtest measures the student's achievement in the areas of history, geography, political science, and economics. Also assessed is the student's ability to apply that knowledge and analyze new information. Darcie's score is in the Average range for the grade. Ask your student to draw simple maps. Share and discuss daily news headlines. Encourage your student to read and discuss content in informational, biographical, or historical fiction books.

STANFORD LEVEL FORM: INTERMEDIATE 4-6
2002 NORMS: Spring National

Scores based on normative data copyright © 2003 by Harcourt Assessment, Inc.

COPIES
PROD.



For Parents:
This report is also
available in Spanish.

Figura 4. Ejemplo de informe de resultados individuales dirigido a padres de familia

Fuente: Stanford Achievement Test series, Tenth Edition. Reporte elaborado en 2003 por Harcourt Assessment, Inc.

INFORMES GRUPALES

A continuación se explica e ilustra un ejemplo de una de las diversas formas en que se puede presentar un informe de resultados grupal. En la figura 5, se encuentran los resultados de la ejecución de un grupo de estudiantes, en donde se puede observar que hay cuatro secciones principales, diseñadas para responder a las siguientes preguntas:

1. **¿Cuál es el puntaje promedio del grupo en el área curricular evaluada?** En esta sección se encuentra una breve explicación del significado de la puntuación global de la ejecución grupal, así como una gráfica que apoya dicha explicación.
2. **¿Están los estudiantes evaluados cumpliendo con los estándares estatales?** Aquí se dividió a los estudiantes en cuatro grupos para realizar la interpretación de los resultados: los que obtuvieron puntuaciones que los ubicaron en un nivel básico, los que obtuvieron resultados en un nivel eficiente, los que fueron ubicados en un nivel de ejecución acelerada, y los de ejecución de tipo avanzada. Se explica en forma concisa lo que cada nivel significa y el rango de puntuaciones que se utilizó para clasificar dichos niveles, y se interpretan las puntuaciones grupales.
3. **¿Cuáles son los siguientes pasos y qué recomendaciones se sugieren?** En esta sección se brindan sugerencias para mejorar el desempeño y apoyar el desarrollo académico del grupo.
4. **¿Qué ejecución tuvo el grupo evaluado en relación a cada estándar de contenido?** En este ejemplo se explican en forma abreviada los resultados grupales en cada una de las áreas curriculares que evaluó la prueba,

destacando lo que el grupo puede hacer (competencias), y no lo que no puede hacer. Se hace mención de las preguntas específicas que fueron utilizadas para evaluar cada área del currículum, así como del porcentaje de respuestas correctas que obtuvo el grupo en cada una de ellas. Con esta información y la que se tiene de la escuela, distrito escolar y estado, se presenta una gráfica de niveles de dominio alcanzados por el grupo, en comparación con los resultados del ciclo escolar anterior, de toda la escuela, del distrito, así como los del estado.

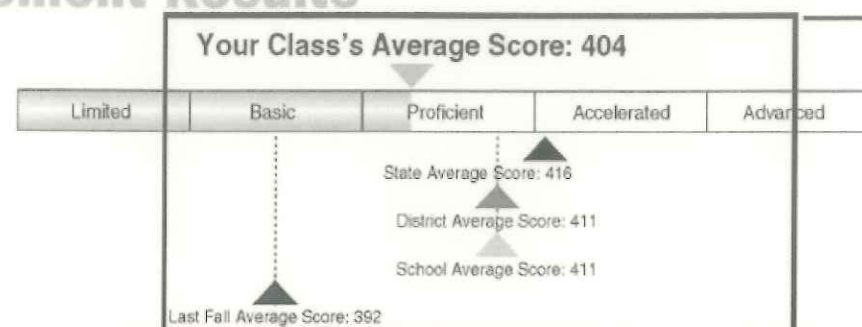


What Is My Class's Average Reading Score?

Reading Achievement Results

CLASS AVERAGE SCORE

Your class's average score is 404. This is lower than the performance of the third-graders in your school, lower than that of third-graders in your district and lower than that of third-graders statewide.



The scale scores in this report are used to keep the performance standards (e.g., the proficient level) comparable from year to year even though a new version of the test is used during each testing cycle. Scores in different subjects cannot be compared.

Are Students Meeting State Standards?

HOW YOUR CLASS PERFORMED

This chart compares the percentage of students from your class who scored in each performance level with the percentage of students who scored in each performance level in your school, your district and the state. (Your class is included in the averages for the other groups.)

Students whose scores fall within the proficient range meet state standards. When compared with students throughout the state, a lower percentage of your students are proficient or above.

Note: This report is not intended for teacher evaluations.

*Fall performance score includes 21 of your students; spring performance score includes 21 of your students.

Recommended Next Steps

Your class's performance on Reading does not suggest a relative weakness compared with other classes in your school. However, to support students' growth, you could target Reading Process with activities such as the ones listed below.

- Have students read aloud appropriate-level texts, pausing to ask questions, make predictions and check comprehension.
- Have students use teacher-created reading guides to help them use text features and identify central ideas (guides can also include diagrams for predictions, summaries or comparisons).
- Divide students into mixed-ability groups. Using the jigsaw method, have each group read part of a text. Have the groups report summaries, predictions and conclusions.

WHAT THE PERFORMANCE LEVELS MEAN

Advanced (432-520)

Students go beyond understanding what they read to explain what they like or do not like about a story.

Accelerated (415-431)

Students understand what they read (e.g., fairy tales, folk tales, poetry) and compare stories with other stories.

Proficient (400-414)

Students understand what they read. They try to list important ideas in the texts they read.

Basic (385-399)

Students understand some of what they read. They use strategies (e.g., clues in sentences and paragraphs, knowledge of word parts) to learn new words.

Limited (250-384)

Students may struggle with simple reading tasks (e.g., fairy tales, folk tales, poetry).



How Did My Class Do on Each Content Standard?

The test has enough questions to report only three levels of performance for each content standard.

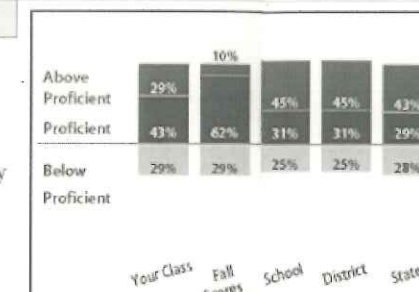
NOTE: The released questions in the tables to the right reflect the composition of the test. The bar graphs directly below represent your class's performance on all the questions. Individual test questions give an incomplete picture of student performance, so it is better to focus on the performance data for each standard.

Starting in July, you can see some of the actual test questions at www.success.ode.state.oh.us

Below you will see your class's performance on the released questions in each content standard.

Acquisition of Vocabulary

Students figure out what new words mean by looking at the words and sentences around the new words. They correctly read common words (e.g., *ant*, *leaf*, *pillow*). They use what they know about parts of words (e.g., *can't*, *lovely*, *helpful*) to understand whole words.

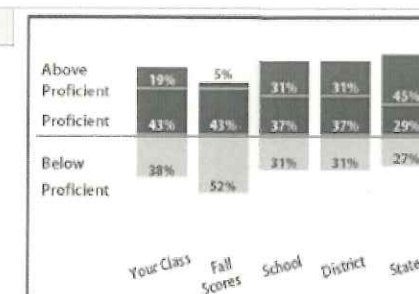


MULTIPLE-CHOICE QUESTIONS

QUESTION NUMBER	2	3	4	8	26
PERCENT CORRECT	100	95	90	86	81

Reading Process

Students read books by themselves and use reading strategies to help understand what they read. For example, they summarize stories, articles or book chapters by putting information in the right order and by including main ideas and important details.



MULTIPLE-CHOICE QUESTIONS

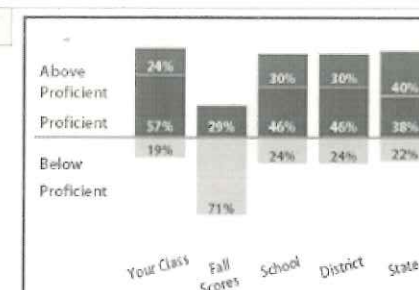
QUESTION NUMBER	10	11	12	21
PERCENT CORRECT	90	95	82	81

CONSTRUCTED-RESPONSE QUESTIONS

QUESTION NUMBER	13	25
AVERAGE POINTS	1.2	1.8
POSSIBLE POINTS	2	4

Informational Text

Students use text features (e.g., titles, index, pictures) to understand informational texts (e.g., maps, instructions, non-fiction). They know how to draw conclusions from information in charts, graphs and diagrams.



MULTIPLE-CHOICE QUESTIONS

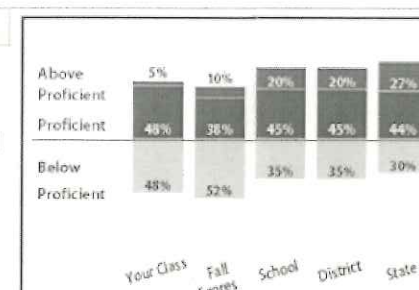
QUESTION NUMBER	7	14	15
PERCENT CORRECT	86	67	76

CONSTRUCTED-RESPONSE QUESTIONS

QUESTION NUMBER	9
AVERAGE POINTS	3.2
POSSIBLE POINTS	4

Literary Text

Students compare and contrast the plots of different stories. They use details from stories to describe characters and setting and can retell what happens. They also explain the features of different types of literature (e.g., fairy tales, folk tales, poems).



MULTIPLE-CHOICE QUESTIONS

QUESTION NUMBER	22	23	24
PERCENT CORRECT	48	82	71

Figura 5. Ejemplo de informe de resultados grupales

Fuente: Departamento de Educación de Ohio (2005). Muestra de "Reporte de resultados grupales" de la Prueba de Aprovechamiento en Lectura, de Ohio.

Anexo 4

Cuestionario para la evaluación de la operación del módulo.

Anexo 4. CUESTIONARIO PARA LA EVALUACIÓN DE LA OPERACIÓN DEL MÓDULO.

Por favor conteste este cuestionario. La información que usted proporcione apoyará la mejora del presente módulo.

En las siguientes secciones se encuentran diferentes elementos que integraron el módulo para el análisis y reporte de resultados del examen, cuya operación se pretende evaluar con este instrumento. La idea es identificar en qué medida se logró cumplir en cada caso con el objetivo propuesto. Para lograrlo le solicitamos que indique en cada inciso, marcando una X dentro del recuadro, del número que refleje su opinión.

1. Instructor

	Totalmente de acuerdo	De acuerdo	Indiferente	En desacuerdo	Totalmente en desacuerdo
a) No llevó a cabo una adecuada moderación de las intervenciones de los participantes.	5	4	3	2	1
b) Cuando hubo dudas proporcionó una realimentación que mejoró la comprensión de los temas.	5	4	3	2	1
c) Realizó ajustes a las expectativas de los participantes de acuerdo a los objetivos y contenidos del curso.	5	4	3	2	1
d) Los ejemplos utilizados durante la instrucción no fueron significativos ni congruentes con cada tema tratado, ni con el nivel del grupo.	5	4	3	2	1
e) Verificó la comprensión de los temas expuestos, a través de preguntas, dinámicas o ejercicios.	5	4	3	2	1
f) Motivó y mantuvo la participación del grupo a través de actividades tendientes a incrementar el interés de los asistentes.	5	4	3	2	1
g) No demostró dominio del tema.	5	4	3	2	1
h) Su exposición no fue clara.	5	4	3	2	1

2. Manual

- a) La información contenida en el manual facilitó la tarea de analizar y reportar resultados de pruebas criterioles de gran escala.
- b) La redacción del texto no es clara ni significativa.
- c) Los temas y subtemas son congruentes con los objetivos de aprendizaje.
- d) Los ejemplos presentados no reflejan la aplicación de la información desarrollada en el contenido de cada tema.
- e) La introducción facilitó la comprensión del contenido del manual.
- f) Se explican con suficiente detalle los procedimientos a realizar.
- g) Los ejemplos, ejercicios y actividades presentados no fueron suficientes para entender lo que hay que hacer.

Totalmente de acuerdo	De acuerdo	Indiferente	En desacuerdo	Totalmente en desacuerdo
5	4	3	2	1
5	4	3	2	1
5	4	3	2	1
5	4	3	2	1
5	4	3	2	1
5	4	3	2	1
5	4	3	2	1

3. Operación del módulo

- a) El tiempo dedicado a cada tema, ejercicio o actividad es suficiente para lograr el objetivo de aprendizaje.
- b) Las actividades realizadas durante el curso no permitieron que se lograra un aprendizaje significativo.
- c) La presentación con recursos de multimedia apoyó la comprensión de conceptos y procedimientos.
- d) No quedó una idea clara sobre la intención del módulo de capacitación.

Totalmente de acuerdo	De acuerdo	Indiferente	En desacuerdo	Totalmente en desacuerdo
5	4	3	2	1
5	4	3	2	1
5	4	3	2	1
5	4	3	2	1

4. Aspectos generales

a) ¿Después de haber cursado el módulo, considera que se encuentra capacitado para realizar de manera independiente los análisis de datos de un examen criterial de gran escala, así como para elaborar los correspondientes informes de resultados?

b) ¿Cuáles fueron las limitaciones más significativas que identificó de este módulo?

¿Qué tan importantes fueron como para limitar su capacidad para realizar los análisis e informes de resultados?

c) ¿Qué cambios sugiere realizar al módulo a fin de hacerlo más útil para analizar los resultados del examen o elaborar los correspondientes informes?

Gracias por su colaboración

Anexo 5

Justificación del cuestionario para evaluar el módulo de capacitación y asesoría para apoyar el análisis, reporte e interpretación de resultados derivados de la aplicación de pruebas criteriosales de gran escala.

Anexo 5. JUSTIFICACIÓN DEL CUESTIONARIO PARA EVALUAR EL MÓDULO DE CAPACITACIÓN Y ASESORÍA PARA APOYAR EL ANÁLISIS, REPORTE E INTERPRETACIÓN DE RESULTADOS DERIVADOS DE LA APLICACIÓN DE PRUEBAS CRITERIALES DE GRAN ESCALA

Para poder llevar a cabo una adecuada evaluación de la operación del módulo de capacitación y asesoría para el análisis y reporte de resultados del examen, es necesario considerar los elementos conceptuales y metodológicos básicos utilizados durante la instrucción, de tal manera que los instrumentos empleados para la evaluación sean congruentes con ellos. En este sentido, si se busca que el curso de capacitación se apegue a estándares de calidad, deberá abordar elementos tales como su diseño (tiempo, material didáctico), congruencia de los objetivos con el tipo de población a quien va dirigido el módulo, las condiciones de operación y el nivel de ejecución; relación objetivos – nivel de aprendizaje; congruencia de las técnicas instruccionales con las características y la cantidad de participantes; entre otros. En la tabla 1, que se presenta al final de este anexo, se observa una síntesis de los factores evaluativos considerados, sus indicadores y variables.

Para efectuar las acciones anteriormente descritas se aplicará un instrumento que se describe a continuación, en donde se evalúa la calidad y estilo del instructor, la calidad y contenido del manual, así como la operación del módulo.

EVALUACIÓN DEL INSTRUCTOR.

En primer lugar, se diseñó una sección dentro del cuestionario cuya función general es recopilar las opiniones de los usuarios respecto a la forma en que el

instructor llevó a cabo la impartición del curso. Este factor evaluativo es de suma importancia, ya que el hecho de contar con un instructor capacitado para la impartición de este módulo, que domine los temas, tenga experiencia como capacitador, y cuyo estilo instruccional permita mantener interés y motivación en los participantes, facilitará el establecimiento de un clima de trabajo y compromiso con el cual se lograrán los objetivos del módulo. Para evaluar al instructor se consideraron dos indicadores.

El primero es el conocimiento y dominio temático que tiene respecto al área de evaluación del aprendizaje; particularmente respecto al análisis e interpretación de resultados, así como la respectiva elaboración de informes de resultados a distintos destinatarios. En este sentido, la variable a considerar es la forma en que llevó a cabo el desarrollo de los diversos contenidos que se van abordando a lo largo del módulo; es decir, la manera en que establece una relación de congruencia entre los temas y los objetivos.

El segundo indicador contemplado dentro del cuestionario es su estilo de comunicación y didáctica. Dentro de este indicador se encuentra la variable del estilo instruccional, en donde se busca determinar en qué medida el instructor llevó a cabo la moderación de las intervenciones de los participantes durante la revisión de los contenidos; asimismo, se busca saber si los participantes consideran si el instructor orientó adecuadamente sus preguntas, y si las dirigió hacia el objetivo de aprendizaje planteado para cada tema. Otra variable contemplada dentro del estilo instruccional es la capacidad que tiene el instructor de adaptarse a las expectativas de los participantes, realizando ajustes según transcurran los objetivos y contenidos del

curso, pero manteniendo la trayectoria conceptual del curso. Una tercer variable evaluada en el cuestionario es la del uso que hace el instructor de los ejemplos; ya sea en forma de ejemplos útiles y apropiados a los temas; o técnicas, ejercicios, dinámicas y preguntas empleadas para esclarecer los aspectos teóricos abordados en cada tema. Finalmente, la cuarta variable considerada en el indicador de comunicación y didáctica es la de claridad, con la cual se busca conocer en qué medida el participante considera que el instructor fue conciso en sus explicaciones y claro en su instrucción, de tal forma que no se presentara ningún problema de comprensión por parte del participante.

EVALUACIÓN DEL MANUAL.

El propósito de diseñar una sección para la evaluación del manual del usuario del módulo fue contar con información respecto a la funcionalidad del documento. Se consideraron dos indicadores principales: contenido y diseño del manual.

En el primer indicador se evalúan variables tales como la información contenida en el manual; es decir, si dicha información se considera útil y actualizada en el campo de la evaluación del aprendizaje. Otra variable contemplada en el indicador de contenido es la estructura que tiene el documento; es decir, si está elaborado de tal forma que desde el momento en que el lector revisa su introducción, se le facilita la comprensión del contenido que revisará en cada una de sus secciones. En la variable denominada metodología se pretende medir el grado en que el participante del módulo considera que dentro del manual se explicaron con suficiente detalle los procedimientos a realizar en cada uno de los temas revisados;

así como si los ejemplos, ejercicios y actividades que se presentaron a lo largo del manual fueron suficientes como para que comprendiera con precisión las actividades a realizar. Por último, con la variable llamada relación temática – objetivo de aprendizaje, se desea obtener información respecto a la congruencia que guarda la relación entre los temas y subtemas revisados, con los objetivos de aprendizaje que se establecen en cada unidad dentro del manual. De esta forma se contará con datos respecto a la claridad temática, así como la lógica y correspondencia de un tema con otro.

En el segundo indicador llamado Diseño se evalúan dos variables. La primera aborda aspectos de la redacción utilizada en el manual; es decir, si el lenguaje utilizado es accesible y de fácil comprensión para el lector, y en particular si la redacción del texto expresa en forma clara y significativa el material, de tal forma que se facilite la comprensión del mismo. Es importante considerar este aspecto en la evaluación del manual, debido a que un documento con tecnicismos excesivos vuelve tediosa y pesada la lectura, desviando el interés y motivación del usuario. La segunda variable evaluada a través de este indicador es la de los ejemplos utilizados a lo largo del manual. Con ella se busca establecer si el usuario considera que reflejan o no la aplicación de la información desarrollada en el contenido de cada tema.

EVALUACIÓN DE LA OPERACIÓN DEL CURSO.

Finalmente, dentro del cuestionario de evaluación de la operación del módulo de capacitación, se consideró importante contar con información respecto a la forma

en que se lleva a cabo dicho módulo. Para ello se diseñó una sección dentro del cuestionario, cuya función general es recopilar las opiniones de los usuarios respecto a indicadores tales como los objetivos del curso, su desarrollo, y algunos aspectos respecto a limitaciones percibidas por los participantes y cambios sugeridos por los mismos. En dicha sección se abordan variables tales como la claridad y precisión con la que se definen los objetivos y cómo éstos reflejan la intención del módulo de capacitación; la medida en que el tipo de actividades que se realizaron durante el curso permitieron que se lograra un aprendizaje significativo; asimismo, se consideró importante tener información respecto al uso de recursos de multimedia (proyector de multimedia, proyector de acetatos, etc.) para apoyar la presentación de los temas y facilitar la comprensión de los conceptos y procedimientos revisados. Otra variable contemplada en la evaluación del desarrollo del módulo es la relación existente entre el tiempo dedicado a la instrucción de los temas y el logro de los objetivos que se pretenden lograr con dicha instrucción, de tal manera que se profundice solamente en aquéllos temas en los que se requiere hacerlo.

Por último, se dedica una sección dentro del cuestionario de evaluación en la cual se evalúan tres aspectos respecto a la opinión del participante del curso: el grado en que el participante considera que se encuentra capacitado para realizar de manera independiente los análisis de datos de un examen criterial de gran escala, y para elaborar los correspondientes informes de resultados; las limitaciones más significativas que identificó en el módulo, y qué tan importantes las considera como para limitar su capacidad para realizar los análisis e informes de resultados. Por último, se consideró de gran relevancia incluir un apartado donde el participante

podiera hacer sugerencias en torno a los cambios que se pueden realizar al módulo a fin de hacerlo más útil para analizar los resultados del examen o elaborar los correspondientes informes.

Con esto se pretende evaluar los principales factores que se consideraron de importancia para conocer la forma en que el módulo de capacitación fue operado.

Tabla 1. Factores evaluativos, indicadores, variables y preguntas para evaluar la operación del módulo de capacitación

FACTOR EVALUATIVO	INDICADOR	VARIABLE	PREGUNTAS
INSTRUCTOR	Conocimiento y dominio temático	Dominio temático	■ Demostró dominio del tema ■ Cuando hubo dudas proporcionó una realimentación que mejoró la comprensión de los temas ■ Llevó a cabo una adecuada moderación de las intervenciones de los participantes ■ Verificó la comprensión de los temas expuestos, a través de preguntas, dinámicas o ejercicios. ■ Motivó y mantuvo la participación del grupo a través de actividades tendientes a incrementar el interés de los asistentes.
		Estilo instruccional	■ Realizó ajustes a las expectativas de los participantes de acuerdo a los objetivos y contenidos del curso.
	Comunicación y didáctica	Adaptabilidad	■ Los ejemplos utilizados durante la instrucción fueron significativos y congruentes con cada tema tratado y con el nivel del grupo.
		Uso de ejemplos	■ Fue clara su exposición
		Claridad	■ La información contenida en el manual facilitó la tarea de analizar y reportar resultados de pruebas críticas de gran escala
MANUAL	Contenido	Información	■ La introducción facilitó la comprensión del contenido del manual
		Estructura	■ Se explican con suficiente detalle los procedimientos a realizar
		Metodología	■ Los ejemplos, ejercicios y actividades presentados fueron suficientes para entender lo que hay que hacer
		Relación temática – objetivo de aprendizaje	■ Los temas y subtemas son congruentes con los objetivos de aprendizaje
	Diseño	Redacción	■ La redacción del texto es clara y significativa ■ Los ejemplos presentados reflejan la aplicación de la información desarrollada en el contenido de cada tema.
CURSO	Objetivos	Claridad y precisión	■ Quedó una idea clara sobre la intención del módulo de capacitación
		Tipo de actividades	■ Las actividades realizadas durante el curso permitieron que se lograra un aprendizaje significativo
	Desarrollo	Uso de recursos	■ La presentación con recursos de multimedia apoyó la comprensión de conceptos y procedimientos
		Relación tiempo instruccional – objetivo de aprendizaje	■ El tiempo dedicado a cada tema, ejercicio o actividad es suficiente para lograr el objetivo de aprendizaje
	Opinión de los participantes	Nivel de competencia	■ Después de haber cursado el módulo, considera que se encuentra capacitado para realizar de manera independiente los análisis de datos de un examen criterial de gran escala, así como para elaborar los correspondientes informes de resultados?
		Limitaciones del curso	■ ¿Cuáles fueron las limitaciones más significativas que identificó de este módulo? ■ ¿Qué tan importantes fueron como para limitar su capacidad para elaborar los análisis e informes de resultados?
		Cambios sugeridos	■ ¿Qué cambios sugiere realizar al módulo a fin de hacerlo más útil para analizar los resultados del examen o elaborar los correspondientes informes?

Anexo 6

**Formato para cotejar los componentes del análisis de la
ejecución de los examinandos**

Curso-taller para la elaboración de exámenes criterios.
Módulo 5. Análisis, reporte e interpretación de resultados de pruebas criterios
de gran escala

Anexo 6. FORMATO PARA COTEJAR LOS COMPONENTES DEL ANÁLISIS DE LA EJECUCIÓN DE LOS EXAMINANDOS

COMPONENTE	SI	NO
1. El análisis de reactivos fue realizado con datos empíricos derivados de la aplicación a gran escala de los modelos de examen?		
2. La base de datos es una tabla de doble entrada?		
La tabla de doble entrada incluye datos sobre:		
▪ Examinandos		
▪ Ítems		
▪ Datos crudos		
▪ Calificación de la prueba		
▪ Puntaje total		
▪ Índice de dificultad		
▪ Índice de discriminación		
▪ Coeficiente de discriminación (r_{bis})		
▪ Índice de confiabilidad de modelos mediante Coeficiente alfa o Kuder-Richardson 20		
▪ Análisis de las respuestas a los distractores muestra los que no fueron seleccionados por al menos un mínimo de los examinandos menos capaces?		
3. Establece relaciones entre variables de contexto y la ejecución del examinando?		
Variables contempladas:		
▪ Variables propias del alumno		
▪ Variables propias de la Escuela		
▪ Variables propias del hogar		
▪ Variables propias de la operación del currículo		
4. Establece estándares de ejecución y puntos de corte?		
5. El análisis de la ejecución de los examinandos está basado en el análisis curricular (retícula)?		
6. Establece niveles de dominio del contenido, por área curricular y/o tipo de organización curricular de que se trate (unidad-tema-subtema, eje-sub eje-línea de formación, o cualquier otro)?		
7. ¿El análisis realizado desagrega la información según los destinatarios?		
8. ¿Establece una desagregación de resultados por zona escolar?		
▪ ¿Analiza los resultados en forma separada por cada zona dentro del municipio y estado?		
▪ ¿Analiza los resultados en forma separada por cada escuela dentro de la zona?		
▪ ¿Hace mención de los factores asociados al desempeño de los estudiantes?		
▪ ¿Resalta los aspectos más interesantes de la ejecución de los alumnos?		
▪ ¿Proporciona información útil a los directores de planteles y maestros de la zona para que realicen la planeación educativa?		

COMPONENTE	SI	NO
¿Proporciona información útil a las autoridades educativas para efectuar revisiones de las políticas educativas? ¿Proporciona información útil para difundir los resultados entre la población de la zona? ¿Analiza los resultados de la evaluación de los alumnos de cada grupo? ¿Está dedicada una parte del análisis a la ejecución general del grupo en lo que respecta a la materia evaluada? ¿Analiza la ejecución de los alumnos en relación a los reactivos que evalúan dominios específicos del conocimiento general que evalúa la prueba? ¿Incluye un análisis categórico de los porcentajes obtenidos por los examinandos con sus puntuaciones, agrupándolos por variable de contexto (edad, género, turno, etc.)? ¿Incluye un análisis categórico de los porcentajes obtenidos por los examinandos con sus puntuaciones, según la escala que se esté utilizando (p. ej. excelente, bueno, regular, malo)? ¿Incluye en el análisis la obtención de los puntajes promedio, por variable? ¿Incluye el uso de gráficas para presentar la información en rangos percentilares, facilitando la comparación de la distribución de los puntajes de diversos grupos? ¿Incluye un análisis estadístico de la variación de los puntajes promedio generada por un muestreo aleatorio y por una prueba poco confiable?		
9. ¿Establece una desagregación de resultados por escuela?		
¿Incluye una distribución de frecuencias de los examinandos, por ciclo escolar y variables de contexto utilizadas (edad, grupo, género, nivel socioeconómico, turno, etc.)? ¿Incluye puntajes promedio y desviaciones estándar de todos los examinandos, de cada ciclo escolar evaluado? ¿Incluye puntajes promedio y desviaciones estándar de los examinandos agrupados por variables de contexto, de cada ciclo escolar evaluado? ¿Incluye una distribución de porcentajes de estudiantes ubicados en cada nivel de dominio, por ciclo escolar y variables de contexto-ciclo escolar? ¿Incluye intervalos de confianza de los puntajes promedio globales de cada ciclo escolar evaluado? ¿Incluye intervalos de confianza de los puntajes promedio globales, por variables de contexto? ¿Incluye una prueba de hipótesis de las diferencias entre los puntajes promedio de los ciclos escolares que se desean comparar?		
10. ¿Establece una desagregación de resultados del estado?		
¿Muestra las tendencias de ejecución de los estudiantes en la cantidad de ciclos escolares establecidos? ¿Proporciona información para saber si son estadísticamente confiables y académicamente significativos los promedios de ejecución de los estudiantes en evaluaciones de ciclos escolares consecutivos? ¿Realiza un análisis para saber cuáles son los factores que pueden contribuir año tras año a las tendencias de la ejecución de los examinandos?		

Anexo 7

Formato para cotejar los componentes del informe de resultados

Curso-taller para la elaboración de exámenes criterios.
Módulo 5. Análisis, reporte e interpretación de resultados de pruebas criterios
de gran escala

**Anexo 7. FORMATO PARA COTEJAR LOS COMPONENTES DEL INFORME DE
RESULTADOS**

COMPONENTE	SI	NO
1. Describe las metas específicas que persigue el informe de resultados?		
2. Comunica los resultados de la evaluación en forma clara, precisa y significativa a los destinatarios?		
3. Explica la forma en que los resultados individuales y grupales ayudarán a la escuela en el mejoramiento del proceso enseñanza – aprendizaje?		
4. Explica y describe los propósitos que orientaron la desagregación de la información?		
5. Explica y describe los propósitos y técnicas utilizadas al reportar la información?		
6. Explica y describe el contexto dentro del cual se llevó a cabo la evaluación?		
7. Incluye información de:		
• Nombre de la escuela		
• Características del centro educativo		
• Descripción socioeconómica de la ciudad en la que se ubica,		
• Cantidad de alumnos que se atienden,		
• Características socioeconómicas del alumnado		
• Características culturales del alumnado		
• Grados académicos que se imparten		
• Materias o áreas que se evaluaron		
• Otros		
8. Explica y describe la naturaleza de la evaluación (prueba)?		
9. Contextualiza al usuario respecto al tipo de examen que se aplicó?		
10. Contextualiza al usuario respecto a la forma en que fue elaborado el examen (planeación, construcción, aplicación, y análisis de resultados)?		
11. Explica y describe quiénes son los destinatarios de la información?		
• Estudiantes		
• Profesores		
• Autoridades educativas		
• Padres de familia		
• Público en general		
• Otro		
12. Explica y describe qué información se les proporcionará a los destinatarios?		
13. Explica y describe qué usos podrán darle los destinatarios a la información?		
14. Explica y describe un análisis fundado, respecto a las razones por las cuales se deben reportar determinados resultados de la evaluación a estos destinatarios en particular?		
15. Si es destinatario de primer nivel, incluye:		
• Resultados obtenidos respecto a la ejecución a nivel estatal y municipal		
• Niveles de los puntajes de los examinados (alto, medio, bajo; u otra clasificación)		

COMPONENTE	SI	NO
Medidas que pueden adoptarse para examinar los resultados de la evaluación y mejorar el proceso educativo		
Forma en que se elaborarán los informes que serán dirigidos a los padres de familia y público en general		
Distribución de promedios de las escuelas del Estado, en cada área del currículum evaluada		
Distribución de promedios de las escuelas, por municipio, en cada área curricular		
Tendencia del municipio y promedios estatales en cada área curricular, durante los cinco años anteriores		
Resultados en grupos de preguntas que evalúan blancos de aprendizaje claves por área curricular		
Gráficas que comparen simultáneamente los resultados de logro contra acciones educativas determinadas (perfil de profesores, recursos de la escuela, etc)		
16. Si es destinatario de segundo nivel, incluye:		
Promedio de ejecución de estudiantes en grupos de preguntas que evalúan blancos de aprendizaje claves, por área del currículum		
Promedio de ejecución de estudiantes en cada área del currículum		
Resultados de ambos aspectos, por municipio		
Presentación de los datos en comparación al año anterior		
17. Si es destinatario de tercer nivel, incluye:		
Promedio de resultados por escuela dentro de la región de servicio, para cada área del currículum		
Promedio de resultados por grupo, por cada unidad y tema		
Tendencia de la ejecución promedio, durante los cinco años anteriores		
18. Si es destinatario de cuarto nivel, incluye:		
Promedio general y por género, de resultados de los estudiantes de la escuela y por área del currículum evaluada		
Porcentaje de aciertos de cada examinando por área del currículum		
Comparación del promedio general y por género, de los resultados de los estudiantes de la escuela, por área del currículum, contra las correspondientes medias de las escuelas del estado		
19. Describe de forma abreviada y significativa la información obtenida a partir del análisis de los resultados de la ejecución de los examinandos en la prueba?		
Incluye información gráfica?		
Incluye tablas?		
Incluye esquemas?		
Incluye elementos adicionales?		

