

18.09.01
60



**INSTITUTO DE INVESTIGACION
PARA EL MEJORAMIENTO
DE LA EDUCACION COSTARRICENSE**

ESCUELA DE FORMACION DOCENTE

INFORME FINAL

PROYECTO DE INVESTIGACION Nº 724-93-327

"ELABORACION DE UNA PROPUESTA METODOLOGICA PARA LA SELECCION
DE ESTUDIANTES PARA LAS PRUEBAS A NIVEL NACIONAL DE LOS
DIFERENTES PROYECTOS QUE CONTIENE EL PROGRAMA
DE PRUEBAS DEL I.I.M.E.C."

M.Sc. Andrea Vindas Lara

**UNIVERSIDAD DE COSTA RICA
FACULTAD DE EDUCACION**

Febrero, 1999

Informe Final

Proyecto de investigación # 724-95-535.

Elaboración de una propuesta metodológica para la selección de estudiantes para las pruebas a nivel nacional de los diferentes proyectos que contiene el programa de pruebas del IIMEC.

Investigadora principal Andrea Vindas

Carga asignada: 1/4 tiempo

El presente informe se divide en dos partes. La primera es para la Vicerrectoría de Investigación y la segunda es el informe técnico.

I PARTE

1. Antecedentes

1.1. Desde el inicio, el proyecto de investigación tuvo cuatro objetivos:

-Proponer diseños muestrales apropiados para las pruebas que ejecuta el Programa de Pruebas y sus proyectos respectivos de acuerdo con las necesidades de información del Ministerio de Educación Pública y los recursos del IIMEC.

-Seleccionar las muestras según la metodología propuesta, incluyendo búsqueda o construcción de marcos muestrales apropiados.

-Asesorar el procesamiento estadístico de las pruebas diagnósticas desde el punto de vista estadístico.

-Asesorar el análisis apropiado de la información, en forma congruente con el diseño muestral ejecutado y para los niveles de información previstos.

Como se puede observar, los objetivos propuestos son de apoyo para los diferentes proyectos que requieren el empleo de muestras de estudiantes. Al finalizar, se podrá contar con una propuesta metodológica para la selección de muestras de estudiantes en primaria o secundaria, en el nivel nacional.

1.2 Este proyecto nunca ha tenido presupuesto. El apoyo para imprimir los diseños muestrales, para el envío de cartas, para solicitudes de archivos, para suministros

de materiales necesarios y otros han sido financiados por el Programa de Medición y Evaluación -PMEE. De igual manera, el pago a un profesional elaboró el programa de cómputo para el diseño muestral de 1995, estuvo a cargo del PMEE, dada la falta de tiempo del equipo de planta del Instituto.

1.3. La duración del proyecto contempla 3 años, de 1996 a 1998, sin embargo, la colaboración se inició desde marzo de 1995, con las diferentes pruebas que ha ejecutado el PMEE. La mayor parte del trabajo superó en demasía y se concentró fundamentalmente el año 1995 y primer semestre de 1996.

1.3 El vencimiento del convenio UCR-MEP, que daba sustento presupuestario a los proyectos de elaboración de pruebas diagnósticas del PMEE, y la creación del CENE-EDU en 1997, modificaron las expectativas de trabajo, aunque éstas continuaron desarrollándose normalmente hasta finales de 1997 dentro de la Universidad. Esto obviamente modifica la finalización de este proyecto.

2. ¿Qué se ha hecho y cómo se ha hecho ?

2.1 Las actividades que se han desarrollado han sido las necesarias para cumplir con los cuatro objetivos propuestos, los cuales se ejecutaron para cada una de las diferentes pruebas llevadas a cabo por el PMEE, las cuales son:

- Prueba diagnóstica de conocimientos para estudiantes de primer ingreso a la educación general básica, en marzo de 1995.

- Prueba diagnóstica de conocimientos (abril-mayo 1995) en el nivel nacional, en las cuatro materias básicas para tercero, sexto y noveno año. (Respecto a conocimientos adquiridos el año anterior).

- Prueba diagnóstica de conocimientos (octubre-noviembre 1995), en el nivel nacional en las cuatro materias básicas para tercero, sexto y noveno año.

- Prueba diagnóstica de seguimiento a una submuestra de estudiantes de primer ingreso, en noviembre de 1995.

- Prueba diagnóstica de conocimientos para estudiantes de primer ingreso a la educación general básica, en marzo de 1996.

- Prueba diagnóstica de conocimientos (octubre- noviembre 1996) en el nivel nacional, en las cuatro materias básicas para tercero y sexto año.

-Prueba diagnóstica de seguimiento a una submuestra de estudiantes de primer ingreso, en noviembre de 1996.

Para cada una de estas pruebas se generó el diseño muestral. Se buscó un marco muestral apropiado y lo más actualizado posible dentro de la información oficial que posee el Ministerio de Educación Pública. Esta labor se realizó en los tiempos previstos en el cronograma.

Los marcos muestrales han sido suministrados al IIMEC, por primera vez, en diskette. Esta actividad es novedosa y se generó por la necesidad de organizar el marco muestral, de acuerdo con los requerimientos de cada una de las diferentes pruebas, que se aplicarán a las poblaciones a las cuales van dirigidas.

Contar con programas de cómputo que lean la base de datos y la organicen en la forma deseada, era una meta de este proyecto y se logró. Para este fin, existe una serie de procedimientos necesarios, en los cuales el Instituto ya tiene experiencia.

Ahora, no solo se cuenta con el programa de cómputo, sino también se poseen otras opciones más expeditas para organizar la base de datos. En mayo de 1997, se solicitaron al Departamento de Estadística del Ministerio las bases de datos separadas para Preescolar, Primaria y Secundaria, las cuales fueron suministradas y se prueba trabajar desde la hoja electrónica EXCEL posibilitando flexibilidad para que el Instituto pueda suministrar a futuro marcos muestrales organizados en forma diferente a lo requerido en las investigaciones aquí apoyadas.

Contar con un marco muestral con las características deseadas fue un proceso difícil. En el primer intento, hubo que emplear tijeras, goma y calculadora para obtener el producto en la forma deseada. El Ministerio de Educación Pública suministró la información, la cual fue retroalimentada hasta lograr el resultado idóneo, según las necesidades de las pruebas participantes en el proyecto.

Los programas de cómputo para efectuar la lectura y ordenamiento del marco muestral han sido elaborados por el equipo de programadores de la sección de cómputo del PMME, en forma conjunta y coordinada con la investigadora.

Se ha apoyado la labor de tabulaciones, presentación de resultados e interpretación de la información, particularmente con las pruebas de primer ingreso y sus seguimientos. Durante 1997, se apoyó la presentación de resultados al Consejo Superior de Educación y la estructuración y presentación de informes-resumen por escuelas y regiones.

Se ha podido organizar con antelación la estructura que lleva el informe de investigación de cada proyecto participante, junto con las tabulaciones y técnicas estadísticas por emplear, a partir de los términos de referencia y de la consulta al señor Ministro acerca de las mismas. Lo anterior permite que el diseño muestral sea congruente con las necesidades, y no improvisar o solicitar a posteriori clasificaciones no contempladas en el diseño y grupos o estratos a comparar. Esto le da congruencia al diseño y al informe final.

El poder diseñar las siete muestras implicó un proceso de búsqueda de opciones de la mano entre lo estadístico y lo práctico cuya sistematización se presenta en la segunda parte de este informe.

El proceso se ha venido afinando, y para cada diseño muestral hay mayor dominio del tema técnico y sustantivo, por parte de la investigadora principal, de los investigadores, y del equipo de programación y cómputo del PMEE.

Se han compartido experiencias entre el equipo Pruebas Diagnósticas de Conocimientos, en diferentes ocasiones, y particularmente con quien coordina, para mejorar la labor de trabajo de campo, desde las vivencias en las diferentes aplicaciones.

2.2. Adicionalmente a los objetivos propuestos, se preparó material para una sesión de trabajo, con la finalidad de compartir con los coordinadores los resultados de una encuesta aplicada, conjuntamente, con la prueba de abril de 1995. En la encuesta se preguntaba por aspectos personales, familiares y de la institución educativa a la que asiste el estudiante. Se efectuó un análisis de tipo descriptivo para esa sesión, el cual, posteriormente, se comentó. Esta actividad fue importante, ya que permitió conocer parte del contexto o características de las estudiantes, de sus familias y de sus escuelas, todo lo cual se relacionó con su rendimiento académico.

3. Resultados

3.1 Los objetivos originales del proyecto se cumplieron a cabalidad.

3.2 La sistematización de los procedimientos empleados para los diseños muestrales, brindan un material importante para futuras investigaciones en el área y para consulta de estudiantes. Es importante resaltar que estas documentaciones de procedimientos en forma detallada, no se encuentran en la bibliografía clásica disponible y, generalmente, a este aspecto se le trata muy brevemente y en forma superficial, dentro de las investigaciones.

3.3 Se fue generando una "cultura muestral", para el equipo de investigadores y encargados de programación y cómputo, respecto a la necesidad de visualizar la congruencia e importancia de los diferentes pasos del diseño en función de los objetivos, del trabajo de campo, de los resultados y del análisis.

3.4 El IIMEC cuenta, a partir de 1995 y como producto de acciones de este proyecto, con la estructura de una base de datos acerca de la matrícula inicial de todas y cada una de las instituciones educativas de primaria y secundaria en el nivel público, privado y subvencionado, por región educativa, provincia, cantón, distrito y poblado, y para cada nivel educativo. Esto hace que no sea necesario, como en otros años, estar solicitando información o tabulaciones particulares, sino que el mismo Instituto las puede generar.

3.5 Las muestras diseñadas tienden a cumplir con la rigurosidad estadística, partiendo de tamaños muestrales definidos, pero diseñados, sobre todo, a partir de los recursos disponibles, y pensando en un trabajo de campo ágil y apropiado a las características particulares de las instituciones educativas, y dentro del contexto social e histórico existente.

3.6 Se pretende obtener, cada año, la base de datos de matrícula inicial que posee el MEP para apoyar o generar proyectos en esa línea.

Proyección:

Para el curso de Medición de la tercera promoción de la Maestría en Evaluación Educativa, se brindó el marco muestral de la región de San José, para que se efectuara un trabajo de investigación en una muestra de instituciones. Esta colaboración puede ser considerada como proyección del proyecto.

En el mes de junio, la Secretaría General de la Coordinación Educativa y Cultural Centroamericana, le encargó a don José Andrés Masís, de CONARE, la búsqueda de una persona que compartiera una charla con los directores de los departamentos de Estadística de los Ministerios de Educación de Centro América, acerca del sistema de información, opciones de paquetes estadísticos y nuevas posibilidades. La investigadora de este proyecto fue invitada a dar esta charla y la experiencia generada en este proyecto le permitió la posibilidad de compartir las limitaciones, proponer opciones y abrir una vía de comunicación entre el IIMEC, los seis funcionarios centroamericanos y la secretaria centroamericana, con la finalidad de desarrollar una mutua colaboración en el futuro. En los próximos días, se concretarán acciones y posibilidades de un proyecto de investigación, tomando como base el Anuario Centroamericano de Estadística de Educación y el esfuerzo de coordinación generado para la producción del mismo.

Se tiene una buena comunicación y cooperación expedita con el Departamento de Estadística del Ministerio de Educación para brindar las bases de datos. Este ha sido un logro que le permite al IIMEC mayor independencia para su accionar. Las funcionarias de apoyo técnico del IIMEC que han colaborado, consideran que trabajar con estas bases de datos, cada día, se hace más fácil a partir de la experiencia que se tiene hasta ahora.

Este proyecto también ha generado inquietudes acerca de la calidad de la información, lo cual ha motivado la posibilidad de plantear una propuesta de investigación, en la línea de evaluar el sistema de información en Costa Rica, incluso de manera comparativa con el resto de Centro América.

Haber sistematizado la información del proceso de diseños muestrales, es importante para estudiantes de estadística (les servirá de ejemplo), y también para estadísticos o muestristas que en el futuro diseñen muestras para pruebas educativas o nacionales. Usualmente lo que se encuentra en los textos o informes de investigación acerca del diseño muestral no es poco preciso, y se trata de un modo somero, por lo tanto, registrar un proceso en detalle tiene importancia.

II PARTE

INFORME TÉCNICO

Introducción

Consideraciones generales respecto a diseños muestrales, partiendo de tamaños muestrales predeterminados.

Si bien el diseño de la muestra debe contener los criterios estadísticos, (es decir su tamaño debería ser tal que para efectos técnicos cumpliera determinados requisitos de error de muestreo y nivel de confianza), solo no estos son los únicos elementos por considerar, sino que se debe incorporar la parte sustantiva o propia del campo específico y el componente práctico, desde la perspectiva de los aplicadores de las pruebas o de instrumentos de recolección de información. Para cumplir con los requerimientos estadísticos, es necesario contar con la información apropiada y actualizada, para la o las variables principales en el estudio, y así poder efectuar las sustituciones en las fórmulas respectivas, de acuerdo con el tipo de muestreo por emplear. No siempre es fácil disponer de esta información en forma actualizada y para los niveles o estratos de interés, es decir, en el nivel general del país o para las zonas, regiones, subgrupos o estratos por considerar. Por lo tanto, no solamente esto es lo guía ese tamaño de muestra.

La opinión y el intercambio de opiniones entre el experto en el área sustantiva del estudio y el estadístico es fundamental. Para ambos, debe quedar claro y en forma precisa los objetivos que se pretenden lograr y así valorar opciones. El otro aspecto importante es la parte operativa y administrativa del trabajo de campo, la cual depende del contexto particular del país, de la coyuntura que se presenta y de las diversas situaciones. En el caso del Programa de Pruebas Diagnósticas de Conocimientos del IIMEC, el contexto se refiere al aula, a la escuela, los circuitos, a la provincia, el tamaño de la institución, a la ubicación geográfica y a otras características de tipo social del ambiente micro y macro, y de la coyuntura en que se efectuará el estudio, además, a los intereses particulares del Ministro de Educación de turno y dentro del contexto del contrato firmado con los organismos internacionales que financiaron el proyecto. Todos estos factores deben tomarse en cuenta para definir el diseño muestral y concretarlo. Estos factores, conocidos en su mayoría por los investigadores del programa o previsible en alguna medida, son elementos que también restringen y determinan las posibilidades reales de un diseño muestral, que sea viable en el momento en que se ejecutará la investigación.

De ahí que para elaborar los diferentes diseños muestrales dentro de este proyecto, se ha procedido a escuchar primero el criterio de los investigadores respecto a su idea de cómo podrían ejecutar el trabajo de campo, sus restricciones y posibilidades. Se intercambian criterios y se barajan opciones de cómo podría ejecutarse el trabajo de campo. Adicionalmente, se solicitó confirmación de los términos de referencia para respaldar las acciones antes de ejecutarlas. En el anexo se pueden encontrar copias de estas consultas. En este caso, para escoger entre diferentes opciones prevaleció el criterio de viabilidad y simplificación antes que el refinamiento técnico. Refinamiento podría ser, por ejemplo, una selección al azar de niños por nivel, o una selección PPT conociendo la matrícula real a la fecha de la visita a la institución o bien confirmación de datos de que dispone el MEP respecto a matrícula inicial, versus matrícula en el mes en que se efectúa la investigación, lo que es ideal, pero poco realista, dados los recursos disponibles. Por ejemplo, es difícil "muestrear" las escuelas unidocentes. Aunque son parte de la población, están ubicadas en zonas alejadas y, en general, son rurales, con poca matrícula, con un solo maestro, con acceso difícil para investigadores que no son de la región. Todo esto hace difícil "muestrearlas" en un porcentaje similar a las demás instituciones, ya que eleva los costos. Para estos casos, se ha considerado la opción de excluirlas en algunas ocasiones (definidas así explícitamente), o la posibilidad de tomar una muestra de menor tamaño una vez organizado el marco muestral geográficamente; en este caso, estas escuelas se distribuirán en la muestra, mediante selección sistemática, con la finalidad de que queden repartidas geográficamente.

El equipo de aplicadores de las pruebas diagnósticas de conocimientos, estuvo integrado por el coordinador y por tres personas adicionales que se dividen el trabajo, las instituciones y los grupos por visitar.

Las instrucciones para los aplicadores deben ser sencillas y definidas con anterioridad, para generar la menor cantidad de sesgos de selección, y evitar la incertidumbre acerca de la labor en cada escuela, de los cuestionarios por aplicar, de la forma de escoger a los estudiantes y otros detalles. Por ejemplo, si se le dice en la escuela XX pregunte por la matrícula inicial, compárela con este dato que usted lleva según la base de datos del MEP, si es igual siga este procedimiento, si es menor haga este otro procedimiento, luego determine el número de grupos de los niveles de interés, anote cuántos niños hay en cada uno, y luego seleccione al azar $x\%$ de cada grupo, empleando la tabla de números aleatorios, proceda a llamar a una aula a todos los estudiantes seleccionados y, posteriormente aplíqueles este cuestionario o prueba. Este es un ejemplo de situaciones que podrían parecer ideales desde la óptica del estadístico, sin embargo, desde la práctica y el contexto particular no son viables, y lejos de alcanzar éxito complican el trabajo de campo, y se incurrirían posiblemente, en más sesgos, inseguridad de las personas y en problemas administrativos.

Para pruebas de primer ingreso y pruebas diagnósticas de conocimientos, se parte de una premisa fundamental que debe ser subrayada: cómo es que el tamaño de las muestras estaba determinado a priori, en los convenios firmados. De ahí que este punto no es modificable, sino que se parte de él. Si el número de la muestra es apropiado o no, depende del análisis, de los estratos o de los grupos por considerar y de las comparaciones que se desean efectuar, incorporando las precisiones y los niveles de confianza de las estimaciones. El poder ir obteniendo información periódica (año a año) irá dando evidencia acerca de la variabilidad de las diferentes variables: contenidos, objetivos, dominios, constructos u otros. Esto permitirá determinar a posteriori cuán homogéneos o heterogéneos tienden a ser los estratos o grupos que se comparan. Esta información permite acumulación de evidencia para la elección posterior de tamaños de muestras o definición de estratos y grupos de comparación, basada en la experiencia particular del caso costarricense.

Tener definido un tamaño muestral, no es suficiente para determinar lo apropiado o no de ese número, ya que no todo es cantidad, sino esa cantidad. Hay que tener claro el para qué de ese tamaño muestral. Es en forma integral que se deben analizar los elementos para determinar lo apropiado de la muestra particular, incluyendo tamaño y diseño, la representación de determinadas porciones en la población (estratos, grupos), y la selección aleatoria de conglomerados, además de la variabilidad de la variable o variables y las comparaciones por efectuar con los datos, en el contexto; es dentro de éste donde se deben valorar la viabilidad y la oportunidad de llevar a la práctica un determinado diseño y selección muestral.

La labor de este proyecto se ha enfocado hacia la parte de diseño de la muestra, es decir, al manejo del marco muestral y a la determinación de la asignación de la muestra y selección de la misma, sin cuestionar el tamaño de muestra, dado que es un punto político, económico y práctico prefijado y estipulado a priori y con carácter legal.

El diseño muestral debía ajustar el tamaño de la muestra definido por el equipo, las posibilidades prácticas y seleccionar la muestra distribuyéndola a lo largo de la geografía nacional, empleando estratificaciones para lograr y asegurar representación de determinados sectores de la población. También se escogieron grupos completos, es decir, conglomerados en las instituciones educativas. El criterio de estratificar y luego seleccionar conglomerados tiende a equilibrar, ya que la estratificación nos asegura que determinadas porciones de la población con características similares dentro, pero diferentes entre sí, quedan representadas dentro de la muestra. La escogencia de conglomerados completos hace que se tienda a tener mayor homogeneidad dentro de los mismos, pero, a la vez, la estratificación asegura representación de determinados sectores, lo cual, aunado al ordenamiento geográfico del marco muestral (provincia cantón y distrito), colabora y compensa los efectos y la representatividad. Se pretendió dejar establecidos con anterioridad, a partir de los términos de referencia y las cartas del señor Ministro de Educación (ver anexo), los análisis de información que se efectuarían posteriormente y la estructura del informe final, para asegurar que el diseño lo permitiera. A la vez se trataría de que la muestra fuera, en lo posible, proporcional al tamaño del estrato o grupo, con la finalidad de facilitar los cálculos y evitar incluir ponderaciones. El hecho de tener que ponderar, requiere del concurso de un estadístico e implica mayor cuidado y posibilidades de sesgos. El diseño muestral debe responder concordantemente al análisis posterior, por ejemplo, si se desea comparar urbano y rural, se esperarían dos estratos, o bien, se dividiría el marco muestral en esos grupos. Si también se desea controlar el tamaño de la institución, por ejemplo, escuelas pequeñas, medianas y grandes, habría otra categoría o tipo de estratificación que dividiría el marco muestral, y así sucesivamente para cualquier otra clasificación deseada. Este trabajo de estratificación fue posible con el uso de programas de cómputo que ordenaron el marco muestral inicial. Manualmente, sería un trabajo demasiado laborioso que implicaría gran cantidad de tiempo. Además, el programa de cómputo incluía otros datos como división de la matrícula, por grado, y el cálculo de diversos subtotales para los diferentes estratos, lo cual a mano o con calculadora implicaría grandes esfuerzos y tiempo:

El objetivo relacionado con que la muestra sea proporcional al tamaño se cumplió desde el papel, es decir, en el diseño de la muestra. Se previó una pérdida y la imposibilidad de ejecutar o aplicar todas las pruebas en forma similar en todas las instituciones, con base en los datos de matrícula inicial, oficiales del MEP. Estos datos de matrícula inicial no tienen un sistema de control que permita verificarlos y ajustarlos

durante el período lectivo, por lo que podrían tener imprecisiones considerables y no mensurables por ahora.

Adequar la muestra a las necesidades prácticas obedece fundamentalmente a las condiciones del proyecto y al tiempo, ya que diseñar muestras complejas debe responder a que es el grupo de investigadores y aplicadores de las pruebas debe existir una formación apropiada o una cultura de muestras para aceptar muestras más sofisticadas, lo que repercute en un aplicación más compleja, trabajo de campo más lento, además de la capacitación y simulación que serían necesarias para el éxito las mismas. Esta cultura debe construirse y no es posible imponerla o lograrla de un día para otro. Con el pasar del tiempo y si es que el Programa de Pruebas sobrevive, se hace necesario hablar de: efectos de diseño, error de muestreo, nivel de confianza, efectos de conglomeración, y otros aspectos más técnicos que se deben discutir entre el MEP, los investigadores y el estadístico, comprendiendo con claridad su significado e implicaciones tanto. También debe saberse que las decisiones al respecto tienen implicaciones en la práctica, y que podrían hacer más complejo el trabajo de campo y el análisis estadístico por realizar. Otro punto que debe comprenderse es la limitación que surge ante facilidades prácticas, de aspectos ideales, de acuerdo con el estadístico.

Por ejemplo, con las personas que se encargan de la parte de informática, el proceso de haberles solicitado marcos muestrales de una u otra forma: corrigiendo, redefiniendo criterios para la disposición del mismo etc, ha implicado justificar, explicar la necesidad de hacer los cambios o de imprimir determinada información adicional o eliminar otra información que no es necesaria. Este proceso las ha obligado a enfrentarse a situaciones de demanda de información rigurosa, con un determinado orden y formato, pero, a la vez, ha permitido que conozcan de otras áreas diferentes a su formación profesional.

Gracias a este proceso de investigación, el IIMEC cuenta con una base de datos de matrícula por institución, nivel, sector y otra información al respecto. Esto es un logro, pues ya no es necesario solicitar información al MEP, cada vez que se requiera alguna estadística al respecto, sino que se puede procesar directamente en el Instituto. Además, se evitan problemas de comunicación; en diversas oportunidades fue difícil el envío por parte del MEP de la información exacta que se necesitaba, y en el formato deseado. Esto implicaba tener que pedir una y otra vez la información y corregir las imprecisiones.

La posibilidad de efectuar el diseño muestral para una prueba de nivel nacional, genera varias interrogantes como lo son: el tamaño de muestra, el marco muestral a emplear, el diseño muestral, los cálculos de estadísticas y la congruencia y viabilidad entre estos factores y los objetivos del estudio. Acerca de estos aspectos se

deben tomar decisiones. Respecto al marco muestral, el concepto teórico indica que es una lista de las unidades estadísticas sujetas a observación o medición. Aquí se pueden barajar opciones de marcos que van desde las listas de clase hasta algún otro sistema más general, en el que se pierde la identidad particular de cada alumno, pero no por ello su acceso o posibilidad de selección.

En cada campo de investigación, difieren las probabilidades de contar con marcos muestrales actualizados, y con el menor número de inconvenientes para eliminar posibilidades de sesgos. Es necesario conocer del campo sustantivo o propio del cual se investigará y del contexto, para poder decidir por una fuente de información pertinente. Este conocimiento no siempre está en manos del profesional en estadística. Si el estadístico carece del conocimiento, es necesario que se acerque a informantes que conozcan de la estructura administrativa propia, o con experiencia en estudios similares, de modo que puedan poner al estadístico en contacto con la realidad particular. Para Costa Rica, en el campo educativo, hay diversas posibilidades de obtener o construir un marco muestral tomando en cuenta la estructura administrativa como lo es oficinas centrales, direcciones regionales, circuitos, escuela/ colegio y aula, ya que todos ellos poseen información útil, para el propósito de seleccionar una porción de estudiantes que ejecuten una prueba de conocimientos. El factor tiempo es una variable importante de considerar en la toma de decisiones a la hora de optar por determinada fuente de información, para construir o que se suministre un marco muestral.

En Costa Rica, el Ministerio de Educación publica la nómina de instituciones educativas. En ella aparece la lista de entidades de educación preescolar, primaria y secundaria con sus respectivas matrículas iniciales. Son ordenadas por región, circuito y en orden geográfico. La información que se suministra es la matrícula inicial por nivel, y, además, variables correspondientes a la institución, caserío, provincia, cantón, distrito, código y tipo de institución. Esta publicación, por lo general, está disponible a mitad de año, que implica un tiempo de espera y adicionalmente una estructura o publicación rígida, para efectos de organizar un marco muestral a partir de ella. El trabajo de construir un marco muestral a partir de la nómina implicaría una opción, pero lenta, con trabajo de tijeras, goma, papel, es decir una tarea ardua de carpintería.

El IIMEC, tradicionalmente, le solicitaba al Departamento de Estadística del Ministerio, el marco muestral organizado según los requerimientos. Cabe destacar el apoyo que siempre brindó el Ministerio, a pesar de las limitaciones de recursos tanto humanos como de equipo. El marco muestral solicitado para el primer diseño muestral, que apoya este proyecto, no correspondía en su totalidad a lo solicitado. Se tomó la decisión de solicitar la base de datos con que se genera la nómina, con el fin de que el

IIMEC construyera el marco muestral según los requerimientos particulares. Aunque puede parecer simple, la comunicación entre el estadístico y el programador, para obtener el marco muestral con la organización necesaria, no siempre es fácil. Este trabajo implica coordinación, labor conjunta, revisión de comandos, ratificación y pruebas para asegurar que el producto contenga las características deseadas. Implica, también, una confrontación con datos o cifras oficiales del Ministerio, como forma de cotejar y controlar que no falten o sobren instituciones en los diferentes estratos y grupos. Solicitar, en el marco muestral, totales y subtotales para estratos y grupos respectivamente, es un elemento que colabora como otro medio de control, para asegurar la mayor perfección posible del marco muestral, es decir, no tener duplicados ni elementos faltantes y obtener total cobertura. La base de datos suministrado por el Ministerio es de tipo DBF, con 21 variables o campos y 123 caracteres en total. En el archivo de secundaria, las primeras trece variables y la última son de tipo carácter y las restantes numéricas, a saber:

Campo	Nombre	ancho del campo
1	código	4
2	nombre de la institución	25
3	provincia	1
4	cantón	2
5	distrito	2
6	poblado	20
7	sector	1
8	sector2	1
9	sector3	1
10	zona	1
11	dirección regional	2
12	circuito escolar	2
13	rama	2
14	total matrícula sétimo	8
15	total matrícula octavo	8
16	total matrícula noveno	8
17	total matrícula décimo	8
18	total matrícula undécimo	8
19	total matrícula duodécimo	8
20	total matrícula institución	8
21	región	2

La base de datos de educación primaria solo difiere de la de secundaria en que no contiene el campo 13 correspondiente a rama y en su lugar continúa lo correspondiente al campo 14. Como campo 20 aparece la variable tipo lógico, correspondiente a la zona (urbano o rural).

Copia del programa de lectura definitivo para organizar el marco muestral de acuerdo al último diseño, se encuentra en los anexos y en diskette en el IIMEC, junto con la base de datos.

Como el diseño muestral, los estratos y grupos que se representarán varían en las diferentes investigaciones, no necesariamente este programa servirá en el futuro, pero permitirá tenerlo como modelo para reformarlo y tomarlo como guía.

Se desprendió un nuevo objetivo incorporado al proyecto. Esto se hizo pensando en necesidades futuras y en otros posibles usos de la base de datos de matrícula. Este objetivo pretende valorar opciones de paquetes de cómputo para organizar el marco muestral en forma más ágil, evitando la necesidad de contar con programadores. La base de datos de la matrícula de 1997 fue solicitada al Ministerio de Educación, y nos fue proporcionada, por separado, para los niveles de preescolar, primaria y secundaria. El Ministerio posee estos archivos en tipo DBF, y nos los facilitaron para leerlo desde EXCEL. Se efectuó la prueba, y por ser archivos ya separados, no hubo problema de lectura y de organización (sort), con los comandos que contiene este programa. Esta posibilidad de manejo con EXCEL brinda gran comodidad y posibilita al IIMEC para incursionar en análisis anuales, retrospectivos, comparativos o descriptivos, por mencionar algunas opciones, entre las cuales se puede efectuar una labor en el futuro. Es importante resaltar que el Departamento de Estadística del Ministerio no cuenta con los recursos necesarios para efectuar una labor de análisis profundo de datos, por lo que el IIMEC podría definir una área de trabajo en ese sentido.

En cada informe de investigación de cada una de las pruebas, se especifican los detalles del diseño muestral de cada una. La importancia de estos dos capítulos, que se presentan a continuación, reside en especificar la organización del marco muestral y las tablas necesarias para determinar el tamaño de muestra en cada estrato o grupo. Esta información, que se puede considerar de carpintería, es necesario sistematizarla para que sirva de material para cursos o estudiantes de cursos de muestreo, ya que no se encuentra en libros de muestreo. Así mismo, el programa de cómputo que se anexa y la base de datos que se incorpora, corresponde al último diseño muestral, ya que es la más refinada respecto a los requerimientos del estudio.

El marco muestral en todos los diseños muestrales se estructuró con base en la lista de instituciones educativas, la cual fue suministrada por el Ministerio de Educación Pública. Para cada una de las investigaciones se tomaron como base, para el diseño muestral, los términos de referencia del contrato de financiamiento de las investigaciones. A la vez, se solicitó al señor Ministro de Educación por medio de una carta, la ratificación de los términos de referencia, con la finalidad de clarificar los estratos o grupos por representar, y la estructura del informe de investigación y la concordancia entre ambos. La ratificación de los términos de referencia tenía, adicionalmente, la finalidad de definir los estratos y grupos o subdivisiones, para los cuales se requería tener resultados y evitar, a posteriori, solicitudes de cruces o resultados para subdivisiones o grupos no contempladas en el diseño. Por ejemplo, por región del Ministerio y no por las de MIDEPLAN, o bien para circuitos específicos o escuelas unidocentes de determinada región.

Capítulo 2

En este capítulo, se consignan los diseños muestrales de las pruebas de primer ingreso: pruebas de inicio de los años de 1995 y 1996 y sus respectivas muestras para la prueba de seguimiento.

Fundamentalmente, las dos muestras en lo que difieren es que la de 1995 implicó una estratificación para las diecinueve regiones educativas existentes, y la de 1996 se basó en las seis regiones de MIDEPLAN. Por tanto, el programa elaborado en FOX PRO para la lectura y organización del marco muestral implicó estos estratos.

Una de las decisiones del Ministerio de Educación solicitar los resultados de la prueba de 1996, para las regiones de MIDEPLAN.

Para no estratificar el marco muestral con las diecinueve regiones de Ministerio de Educación sino con las seis de MIDEPLAN, fue necesario obtener los decretos de la regionalización actualizados y clasificar las regiones y sus circuitos educativas dentro de ellas. Esto implicaba que en el programa de cómputo, para elaborar el marco muestral, lo primero que se debía hacer era efectuar la clasificación de las instituciones educativas en una de las seis regiones del MIDEPLAN. Para esto fue necesario efectuar la equiparación y así determinar cuales provincias, cantones, distritos y caseríos corresponden a las regiones de MIDEPLAN y generar así una nueva variable con dicha identificación.

Las dos muestras de seguimiento se encaminaron por el mismo procedimiento de ordenación y selección.

3.1 Muestra Prueba de Primer Ingreso 1995 y 1996.

En marzo de 1995, se diseñó la muestra de primer ingreso. El tamaño de muestra establecido fue de 6200, y de acuerdo con ella se ejecutó el diseño muestra. Esa cifra representa un 6 % de la población respectiva (104125). La coordinadora del proyecto, Máster Ana Teresa León Sáenz, de acuerdo con los objetivos del proyecto, definió los niveles (estratos y grupos) necesarios de representar en la muestra. Fundamentalmente, y como primera estratificación se fijaron las diecinueve regiones del MEP. Adicionalmente, interesaba el tamaño de la escuela (unidocente, dirección 1 a 5), localización (urbana, rural) y modalidad (pública, privada, subvencionada). El primer paso fue construir una tabla con la cifra de población estudiantil y con el número de escuelas correspondientes. Este paso es importante para tener idea de la distribución de ellas, y de la posibilidad real de representación de cada subgrupo dentro de los estratos (regiones, en este caso). En algunos casos se unieron grupos fundamentalmente por razones de escasa matrícula. Estas uniones, que repercutirían en el diseño muestral, aparecen en los siguientes cuadros en un recuadro conjunto. Por ejemplo, en el cuadro siguiente la unión se realizó para las instituciones privadas urbanas, las dirección 1 y las unidocentes. Cabe aclarar que el concepto de unidocente se ha aplicado en Costa Rica para escuelas públicas, sin embargo, aquí se utiliza para clasificar, también, a escuelas no públicas pero de escasa población. Una vez generado el marco muestral, se revisaron los totales y subtotales y fueron comparados con diferentes datos oficiales del MEP, con la finalidad de comprobar la concordancia entre lo planeado como marco, el programa de cómputo y los datos oficiales. Esta comparación resultó altamente satisfactoria, ya que los totales y subtotales coincidían. Solo en algunos pocos casos se presentaron diferencias mínimas que fueron consideradas irrelevantes. Es importante mencionar que el programa de cómputo lee la base de datos, no la modifica ni la reordena, por lo tanto, no deja constancia de la lectura ni de la extracción de información. Hay necesidad de asegurarse y comprobar que no falten o no queden excluidas instituciones educativas dentro del marco muestral. Además de la revisión, con el programador, de las indicaciones acerca de lo que se pide que ejecute el programa, la revisión de los listados con totales y subtotales de matrícula con datos oficiales fueron controles empleados que permitieron detectar errores.

Cuadro 1
Región 1. San José. Número de estudiantes
por tipo de escuela y número de escuelas en la muestra:

		Unidoc.	D1	D2	D3	D4	D5	Total n
Pública	Urbano	-	1987	1324	3590	3744	12538	46
	Rural	199						
Privada	Urbano	818		541	1114	298	397	6
	Rural	-	-	-	-	-	-	-
Subv.	Urbano	-	-	114		111	-	1
	Rural	-	-	-	369	-	-	

Total: 27144 estudiantes.

*Muestra: 53 grupos.

Total: 17531 estudiantes.

Muestra: 33 grupos.

Cuadro 2
Región 2. Puriscal. Número de estudiantes
por tipo de escuela y número de escuelas en la muestra:

		Unidoc.	D1	D2	D3	D4	D5	Total n
Pública	Urbano	-	-	-	-	150	183	3
	Rural	304	756	163	-	-	-	
Privada	Urbano	9	-	-	-	-	-	-
	Rural	-	-	-	-	-	-	-
Subv.	Urbano	-	-	-	-	-	-	-
	Rural	-	-	-	-	-	-	-

Total: 1565 estudiantes.

Muestra: 3 grupos.

Total: 1845 estudiantes.

Muestra: 3 grupos.

Cuadro 3
Región 3. Alajuela. Número de estudiantes
por tipo de escuela y número de escuelas en la muestra:

		Unidoc.	D1	D2	D3	D4	D5	Total n
Pública	Urbano	-	-	-	1236	1385	12046	33
	Rural	23	1374	1091				
Privada	Urbano	63	-	161	-	-	-	1
	Rural	-	-		100	-	-	
Subv.	Urbano		20	39	-	-	-	1
	Rural	-	-	-	-	-	-	

Total: 17538 estudiantes.

Muestra: 35 grupos.

Cuadro 4
Región 4. San Ramón. Número de estudiantes
por tipo de escuela y número de escuelas en la muestra:

		Unidoc.	D1	D2	D3	D4	D5	Total n
Pública	Urbano	-				156		6
	Rural	211	1354	580	627	-	837	
Privada	Urbano		21	-	-	-	-	1
	Rural	30	-	-	-	-	-	
Subv.	Urbano	-	-	30	-	-	-	1
	Rural	-	-	-	-	-	-	

Total: 3846 estudiantes.

Muestra: 8 grupos.

Cuadro 5
Región 5. Cartago. Número de estudiantes
por tipo de escuela y número de escuelas en la muestra:

		Unidoc.	D1	D2	D3	D4	D5	Total n
Pública	Urbano	-	-					
	Rural	-	-	1262	1948	975	3232	18
Privada	Urbano	92		-	92	-	-	
	Rural	-	36	-	-	-	-	1
Subv.	Urbano	-	-	-	-	-	-	-
	Rural	-	-	-	-	-	-	-

Total: 9610 estudiantes.

Muestra: 19 grupos.

Muestra: 16 grupos.

Cuadro 6
Región 6. Turrialba. Número de estudiantes
por tipo de escuela y número de escuelas en la muestra:

		Unidoc.	D1	D2	D3	D4	D5	Total n
Pública	Urbano	-	1173	-	848	-	185	3
	Rural	293		92		-	-	
Privada	Urbano	15	32	-	-	-	-	1
	Rural	-	-	-	-	-	-	-
Subv.	Urbano	-	-	-	76	-	-	1
	Rural	-	-	-	-	-	-	-

Total: 2714 estudiantes.

Muestra: 5 grupos (5,339).

Muestra: 3 grupos (5,18).

Cuadro 7
Región 7. Heredia. Número de estudiantes
por tipo de escuela y número de escuelas en la muestra:

		Unidoc.	D1	D2	D3	D4	D5	Total n
Pública	Urbano	-	-	-	1708	1342	2655	14
	Rural	275	1093	676			-	
Privada	Urbano			69	70	-	-	1
	Rural	127	76	-	-	-	-	
Subv.	Urbano	-	-	-	-	-	-	-
	Rural	-	10	-	-	-	-	1

Total: 8101 estudiantes.

Muestra: 16 grupos.

Cuadro 8
Región 8. Liberia. Número de estudiantes
por tipo de escuela y número de escuelas en la muestra:

		Unidoc.	D1	D2	D3	D4	D5	Total n
Pública	Urbano	-				475	372	2
	Rural	183	529	345	488	-	-	
Privada	Urbano	-	-	73	-	-	-	1
	Rural	-	-		-	-	-	
Subv.	Urbano	-	-	-	-	-	-	-
	Rural	-	-	-	-	-	-	-

Total: 1618 estudiantes.

Muestra: 3 grupos (3,18).

Cuadro 9
Región 9. Nicoya. Número de estudiantes
por tipo de escuela y número de escuelas en la muestra:

		Unidoc.	D1	D2	D3	D4	D5	Total n
Pública	Urbano	-	-	90	118	281	-	3
	Rural	495	667			-	-	
Privada	Urbano	-	44	-	-	-	-	1
	Rural	-		-	-	-	-	-
Subv.	Urbano	-	-	-	-	-	-	-
	Rural	-	-	-	-	-	-	-

Total: 1695 estudiantes.

Muestra: 4 grupos (3,33).

Cuadro 10
Región 10. Santa Cruz. Número de estudiantes
por tipo de escuela y número de escuelas en la muestra:

		Unidoc.	D1	D2	D3	D4	D5	Total n
Pública	Urbano	-	-	-		264	-	3
	Rural	155	766	172	365	-	-	
Privada	Urbano	-	-	30	-	-	-	1
	Rural	-	-	-	-	-	-	-
Subv.	Urbano	-	-	-	-	-	-	-
	Rural	-	-	-	-	-	-	-

Cuadro 11

Región 11. Cañas. Número de estudiantes por tipo de escuela y número de escuelas en la muestra:

		Unidoc.	D1	D2	D3	D4	D5	Total n
Pública	Urbano	-	-	-	-	-	-	2
	Rural	349	519	218	321	152	264	
Privada	Urbano	-	9	-	-	-	-	1
	Rural	6	-	-	-	-	-	
Subv.	Urbano	-	-	-	-	-	-	1
	Rural	-	6	-	-	-	-	

Total: 1844 estudiantes.

Muestra: 4 grupos.

Cuadro 12

Región 12. Puntarenas. Número de estudiantes por tipo de escuela y número de escuelas en la muestra:

		Unidoc.	D1	D2	D3	D4	D5	Total n
Pública	Urbano	-	-	-	-	-	-	7
	Rural	554	1185	439	674	460	1009	
Privada	Urbano	-	-	-	-	-	-	1
	Rural	20	-	108	-	-	-	
Subv.	Urbano	-	-	-	114	-	-	1
	Rural	-	-	-	-	-	-	

Total: 4563 estudiantes.

Muestra: 9 grupos.

Cuadro 13
Región 13. Aguirre. Número de estudiantes
por tipo de escuela y número de escuelas en la muestra:

		Unidoc.	D1	D2	D3	D4	D5	Total n
Pública	Urbano	-			-	-	141	3
	Rural	263	437	99	-	72	-	
Privada	Urbano	-	-	-	-	-	-	-
	Rural	-	-	-	-	-	-	-
Subv.	Urbano	11	-	-	-	-	-	1
	Rural	-	-	-	-	-	-	-

Total: 1023 estudiantes.

Cuadro 14
Región 14. Pérez Zeledón. Número de estudiantes
por tipo de escuela y número de escuelas en la muestra:

		Unidoc.	D1	D2	D3	D4	D5	Total n
Pública	Urbano	-	-	-			164	9
	Rural	1166	2555	506	699	346	-	
Privada	Urbano	17	21	18	-	-	-	1
	Rural		-	-	-	-	-	-
Subv.	Urbano	-	-	-	67	-	-	1
	Rural	-	-	-	-	-	-	-

Cuadro 15
Región 15. Coto. Número de estudiantes
por tipo de escuela y número de escuelas en la muestra:

		Unidoc.	D1	D2	D3	D4	D5	Total n
Pública	Urbano	-	-	-	-	-	-	9
	Rural	1212	2674	493	763	150	-	
Privada	Urbano	-	15	-	-	-	-	1
	Rural	10	-	-	-	-	-	
Subv.	Urbano	-	-	-	-	-	-	-
	Rural	-	-	-	-	-	-	-

Cuadro 16
Región 16. Limón. Número de estudiantes
por tipo de escuela y número de escuelas en la muestra:

		Unidoc.	D1	D2	D3	D4	D5	Total n
Pública	Urbano	-	-	-	-	-	-	11
	Rural	366	3161	1268	793	541	994	
Privada	Urbano	-	40	122	-	-	-	1
	Rural	4	-	-	97	-	-	
Subv.	Urbano	-	-	59	-	-	-	1
	Rural	-	-	-	-	-	-	-

Total: 7445 estudiantes.

Muestra: 13 grupos.

Cuadro 17
Región 17. Guápiles. Número de estudiantes
por tipo de escuela y número de escuelas en la muestra:

		Unidoc.	D1	D2	D3	D4	D5	Total n
Pública	Urbano	-	-	-	-	-	-	8
	Rural	214	1698	523	798	394	599	
Privada	Urbano	15	-	-	-	-	-	1
	Rural	-	43	-	-	-	-	
Subv.	Urbano	-	-	-	-	-	-	-
	Rural	-	-	-	-	-	-	

Total: 4301 estudiantes.

Muestra: 8,46 grupos = 9 grupos.

UNIDAD DE INFORMACION Y REFERENCIA
 21 NOV 2001
 IIMEC

18-09-01
 60

1010 0345

Cuadro 18
Región 18. San Carlos. Número de estudiantes
por tipo de escuela y número de escuelas en la muestra:

		Unidoc.	D1	D2	D3	D4	D5	Total n
Pública	Urbano	-	-	-	-	-	-	-
	Rural	1007	2690	391	792	288	210	9
Privada	Urbano	-	-	-	-	-	-	-
	Rural	35	-	-	-	-	-	1
Subv.	Urbano	-	-	-	-	-	-	-
	Rural	-	-	-	76	-	-	1

5489 estudiantes.

Muestra: 11 grupos.

Cuadro 19
Región 19. Upala. Número de estudiantes
por tipo de escuela y número de escuelas en la muestra:

		Unidoc.	D1	D2	D3	D4	D5	Total n
Pública	Urbano	-	-	-	-	-	-	-
	Rural	491	1557	46	179	-	-	5
Privada	Urbano	-	-	-	-	-	-	-
	Rural	-	-	-	-	-	-	-
Subv.	Urbano	-	-	-	-	-	-	-
	Rural	-	-	-	-	-	-	-

Total: 2273 estudiantes.

Muestra: 4,47 grupos = 5.

En resumen, para cada región, la muestra, el número de escuelas, quedó de la siguiente forma:

San José 53, Puriscal 3, Alajuela 35, San Ramón 8, Cartago 19, Turrialba 5, Heredia 18, Liberia 3, Nicoya 4, Santa Cruz 4, Cañas 4, Puntarenas 9, Aguirre 3, Pérez Zeledón 11, Coto 11, Limón 15, Guápiles 9, San Carlos 11 y Upala 5. Para un total de 228 instituciones.

Por razones de tiempo y de recursos, se excluyeron de la selección las escuelas unidocentes y aquellas que tenían en primer año diez niños o menos.

En cada región, el tamaño de muestra n , en los cuadros anteriores, significa la cantidad de grupos de primer grado por seleccionar. La selección se efectuó en forma sistemática, en cada renglón, de modo que la misma se distribuía a lo largo de los diferentes tamaños de instituciones.

Para efectuar la selección, se hacía necesario tener un marco muestral ordenado en la forma en que se encuentran los cuadros anteriores. De esta manera, las instituciones estaban en diecinueve listados diferentes, cada listado ordenado en seis subgrupos: público urbano, público rural, privado urbano, privado rural, subvencionado urbano y subvencionado rural. En cada uno de ellos, a su vez, las instituciones ordenadas ascendientemente por matrícula, es decir por su tamaño. La selección sistemática permitió distribuir la muestra entre los diferentes tamaños de escuelas.

3.2 Muestra Prueba de seguimiento de Primer Ingreso 1995

El diseño de esta muestra se justifica en la necesidad de dar seguimiento a una sub-muestra de la seleccionada a inicio del año, con la finalidad de medir su progreso en las diferentes áreas valoradas en el instrumento aplicado. Esto equivale a aplicar una prueba a inicio del año y la misma a final del año.

La prueba de seguimiento tenía como objetivo determinado y convenido así con el MFP, aplicar una prueba a un 10 % de la muestra evaluada a inicio de año, lo cual implicaba aplicar 930 pruebas, de las cuales, por motivo de la huelga de maestros, se hicieron efectivas solamente 345. Esto implica un 37 % de la muestra prevista.

Para la selección de la muestra para la prueba de seguimiento, se dividió la muestra inicial de la siguiente forma: se clasificó cada institución educativa, alguno de cuatro estratos o regiones, según la zona geográfica de pertenencia, identificada por la subregión educativa, a saber:

Estrato o Región 1: San José

Estrato o Región 2: Cartago, Alajuela y Heredia.

Estrato o Región 3: Puriscal, San Ramón, Turrialba y Pérez Zeledón.

Estrato o Región 4: Nicoya, Liberia, Cañas, Santa Cruz, Puntarenas, Coto, Upala, Limón, Guápiles, San Carlos y Aguirre.

El marco muestral lo constituyó la lista de la muestras de escuelas visitadas en el mes de abril de 1995. Este marco muestral requirió la construcción de cuatro listas separadas, una para cada estrato. El diseño implicaba seleccionar una muestra estratificada de instituciones, con probabilidad proporcional al tamaño (PPT) del estrato. El tamaño del estrato determinado por el número de estudiantes.

Para ello se empleó la ayuda de una hoja electrónica, en la cual se procedió, para cada Región o Estrato por separado, a lo siguiente:

Destinar tres columnas para:

de conglomerado (lista de instituciones visitadas, con numeración secuencial de uno en adelante)

tamaño del conglomerado (número de estudiantes de la muestra de abril, en esa institución)

total acumulado.

Con base en ello se obtuvo, para cada Estrato o Región :

$N_i = \#$ conglomerados

$T_i =$ suma de los tamaños de los conglomerado

$t_i =$ tamaño de muestra deseado

Se calcula el tamaño promedio por conglomerado = $T_i / N_i = p_{ci}$, es decir, el tamaño de la población del Estrato dividido por el número de conglomerados con que cuenta el estrato.

Cabe mencionar que el concepto población en esta muestra de seguimiento, corresponde a la muestra completa de abril.

$$t_i / p_{ci} = z_i \quad T_i / z_i = k = \text{fracción de muestreo.}$$

Con la fracción de muestreo se selecciona, al azar, un número entre 1 y k, que es el punto de arranque, y luego se continúa en forma sistemática.

	REGIÓN			
	1	2	3	4
Tamaño # conglomerado N_i	58	65	22	49
# población escolar T_i	1662	1715	575	1162
Tamaño muestral deseado (10 %)	166	172	58	116
Población promedio por conglomerado	28,66	26,38	26,14	23,71
$t_i / p_{ci} = z_i$	5,79 = 6	6,52 = 7	2,21 = 3	4,89 = 5
$T_i / z_i = k$	277	245	192	232
# al azar entre 1 y...	100	233	178	041
n obtenida luego de la selección	183	186	80	162

El n obtenido después de la selección se calcula sumando el número de estudiantes de cada conglomerado, de acuerdo con los datos de la muestra de inicio de año, tal y como se registraron en el marco muestral. La mayor diferencia entre el

10 % y el valor real propuesto para la muestra, se presenta en la región 4. Esta región es la que aglutina mayor número de regiones de Ministerio de Educación, por lo que este aumento tiende a distribuirse en más regiones. La variabilidad en el tamaño de los conglomerados es precisamente el factor que determina que el tamaño propuesto y real difieran.

La selección sistemática determinó cuál grupo de niños se incluye para conformar la muestra. En los grupos seleccionados se aplicó la prueba nuevamente, y se compararían sus resultados con los de inicio de año. De esta forma, se determina el avance logrado por los estudiantes.

	ZONAS				TOTAL
	1	2	3	4	
Población.	27467	25486	13517	37655	104125
Matricula inicial					
Porcentaje	26,38	24,48	13,0	36,13	100,0
No. alumnos muestra inicio de año	1662	1715	575	1162	5114
Porcentaje	32,5	33,53	11,24	22,72	100,0
No. de instituciones seleccionadas	58	65	22	49	194

Lo anterior indica que en la zona 4 fue donde se obtuvo menor tasa de respuesta o de acceso a los centros educativos.

El tamaño muestral previsto, para la prueba de inicio de año, era de 6200, sin embargo, solamente 5114 estudiantes fueron evaluados, lo cual equivale al 82,5 %. Este porcentaje no se distribuyó proporcionalmente en cada una de las zonas, como se aprecia en el cuadro anterior. Si se hubiera cumplido que la no respuesta (independientemente del motivo de la misma) hubiera sido proporcional, por zona

geográfica, sería de esperar proporciones iguales, en las dos filas de porcentaje en ese cuadro.

Los datos de la muestra de seguimiento (submuestra) se encuentran a continuación :

	ZONAS				TOTAL
	1	2	3	4	
Tamaño de muestra deseado	167	172	58	117	514
No. de instituciones seleccionadas	6	7	3	5	21
Tamaño de muestra obtenido según la matrícula de las instituciones seleccionadas	183	186	80	162	611
Tamaño real obtenido:					
No.	91	133	46	75	345
%	26,4	38,6	13,3	21,7	100,0

La posibilidad de efectuar cálculos para el error de muestreo o la precisión, con el nivel de confianza dado, implicaría para cada una de las variables investigadas, conocer su medida de variabilidad respectiva y por región, por zona geográfica o por otra subdivisión deseada, lo cual no se tiene. El no contar con la información precisa implica que se puede simular, ya sea para planear la muestra o posteriormente a la obtención de la misma, a partir de los resultados que ella arroja. Además, el muestreo empleado es complejo. Suponiendo que el efecto de la estratificación tiende a compensarse con el efecto de la conglomeración, se podría estimar con la fórmula de muestreo simple al azar sin reemplazo de poblaciones finitas. En este caso la sustitución de la fórmula suponiendo que la variable de interés es la proporción de estudiantes que domina el aspecto o tema, y con un 90 % de confianza, y para la prueba como un todo (no dividida en componentes) .

$$N = 104125 \quad n = 5114 \quad z = 1,645 \quad p = .5 \quad q = .5$$

$$n = n_0 / (1 + n_0 / N)$$

$$n_0 = 5378 \quad n_0 = (z^2 \cdot p \cdot q) / d^2$$

La precisión o el error de muestreo sería de 1,1215 %

Si se calcula con una confianza del 95 %, el valor del error de muestreo sería de 1,336 %

Si se eleva la confianza al 99 %, la precisión puede alcanzar un nivel de 1,59 % en la muestra de inicio de año en el plano nacional.

Cabe no olvidar que en la población de estudio de la evaluación de inicio de año, se excluyeron las escuelas unidocentes y las que tenían como matrícula inicial, en primer año, menos de 10 niños. Si se dispusiera de los datos de la matrícula por región, eliminando estos grupos, unidocentes y primer año con menos de 10 niños, podrían recalcularse los cuadros anteriores, para lograr mayor precisión y poder determinar cuánta representación hay en la muestra, respecto a cada región o zona geográfica como resultado obtenido y no el planeado, ya que el planeado se mantiene en un 6 % de la población, lo cual concuerda con el diseño y planeado, con la población meta (eliminando unidocentes y escuelas dirección 1 con menos de 10 niños en primer año).

Adicionalmente, se había calculado cerca de un 10 % de pérdida o no respuesta, lo cual se ajustó dado que la muestra sería de 220 grupos, y suponiendo un tamaño promedio de los grupos de 30,5 alumnos. Esto arrojaba un total de muestra de 6710 estudiantes, de los cuales, en realidad, se obtuvo solo un 76,2 %. Cabe mencionar que los datos de las matrículas iniciales en las instituciones no fueron constatados en el campo, en el momento de la recolección de información, de ahí que la pérdida pueda deberse a factores, entre otros, como falta de control de los datos reportados ante el MEP, por cada una de las escuelas; deserción; no respuestas; imposibilidad de acceder a la escuela; ausencia de niños el día de la aplicación de la prueba. Estos son los factores explicativos más probables.

El error de muestreo de la submuestra, puede ser visto desde dos perspectivas, una, como muestra independiente, sin embargo, la finalidad de la misma es como posttest, o seguimiento; la otra, es en relación con la muestra de inicio de año. Para el primer caso, el error de muestreo, con un 95 % de confianza, se calcula como la desviación estándar de la prueba completa dividido por la raíz cuadrada del tamaño de

muestra obtenido, y multiplicado por el valor de 1,96 correspondiente a un 95 % de confianza en una curva normal estándar.

$$1,96 (S) / \sqrt{ n }$$

Para el segundo caso, la varianza muestral, para ser sustituida en la fórmula anterior (luego se obtiene su raíz cuadrada), se debe calcular a partir de la información de las dos muestras, la de inicio de año y la de seguimiento, de la siguiente forma: varianza de la diferencia entre la primera muestra menos la segunda muestra, es igual a la varianza de la primera muestra, más la varianza de la segunda muestra, menos dos veces la covarianza de ambas muestras.

La información de la covarianza no está disponible, de ahí que se pueda partir de varios supuestos como, por ejemplo, que la covarianza es cero, lo cual es poco verosímil. Desde este enfoque, se podría, también, suponer que la varianza de la primera muestra es la misma, por lo que el error estándar puede ser calculado como varianza de la primera muestra dividido por el tamaño de esa muestra, más la varianza de la segunda muestra dividido por el tamaño de la segunda muestra. Otra opción es calcular la covarianza, o bien estimarla una vez obtenido el coeficiente de correlación de Pearson, entre el puntaje de la prueba de inicio de año con la de seguimiento. Por ejemplo, se pudiera determinar la correlación en 0,5, este valor se multiplicaría por la varianza de la muestra de inicio de año, esto implica que se supondría que ambas desviaciones estándar (muestra inicial y de seguimiento) son iguales, y el producto de ambas da como resultado la varianza de la primera muestra.

Con la información disponible en el nivel de la prueba como un todo, se dispone de la desviación estándar (24,88) de la prueba de inicio de año, por lo que se puede suponer que esta es igual para inicio y seguimiento.

$1,96 (24,88) / \sqrt{ 345 } = 2,625$ puntos, es el error de muestreo estimado con un 95 % de confianza.

3.3 Muestra Pruebas de Primer Ingreso 1996.

Este diseño muestral se efectúa para la aplicación de la prueba en marzo de 1996. El marco muestral más actualizado, en el momento de la selección, corresponde al de la matrícula de 1995. Por lo tanto, las instituciones nuevas, es decir, creadas a partir de, no tienen posibilidad de ser incluidas en la muestra.

Se excluyeron las escuelas públicas, privadas y subvencionadas que por su matrícula se clasificarían como Dirección 1 (con a lo sumo 10 estudiantes, a lo sumo en primer nivel) o unidocentes.

Según la información oficial, la matrícula de inicio corresponde a 104125 estudiantes en primero grado. El 6 % de ella equivale a 6252 alumnos, que constituirían la muestra.

La investigadora principal del proyecto determinó, con base en los resultados de 1995, que el tamaño promedio de los grupos era de 35 alumnos. Se supone que este dato se mantiene para 1996.

Se elegirá una sección de cada una de las instituciones educativas seleccionadas, lo cual equivale a tomar 179 instituciones, para alcanzar la meta de 6252 alumnos ($179 * 35 = 6252$).

El marco muestral se estructuró de acuerdo con las necesidades del equipo investigador. Los grupos de la población que debían estar representados corresponden a las seis regiones de MIDEPLAN (no a las del MEP). Además, se requería la representación de escuelas públicas, subvencionadas y privadas, clasificadas en pequeñas, (dirección 1) y grandes (dirección 2 hasta 5); también las escuelas clasificadas dentro del programa "Urbano Marginal". Luego de ordenar el marco muestral con el programa de cómputo, según este formato, se obtuvieron los siguientes resultados:

Cuadro 20
Clasificación del marco muestral, por tipo y tamaño.

		# escuelas	# de niños	%	Tamaño de muestra
PRIVADAS	Dirección 1	54	1280	1,34	3
	Dirección 2-5	48	3572	3,73	7
SUBVEN- CIONADAS	Dirección 1	3	34	0,04	1
	Dirección 2-5	14	1005	1,05	3
PÚBLICAS	Urbano marginal	28	4984	5,2	9
	Dirección 1	1396	26686	27,84	50
	Dirección 2-5	572	58283	60,81	109
TOTAL		2115	95844	100,0	182

El tamaño de muestra queda determinado proporcionalmente de acuerdo con el número de alumnos.

$$(95844/104193) * 100 = 92 \%$$

Esta cifra implica que el 8 % de la población es la que queda excluida por corresponder a escuelas pequeñas.

El marco muestral organizado, de acuerdo con el esquema del cuadro anterior y por región, permitió conformar el siguiente cuadro:

Cuadro 21
Cifras del marco muestral por región

	CENTRAL		BRUNCA		H. NORTE		H. ATL.		PAC. CTL.		CHOROT.		TOTAL	
	# inst.	matr.	# inst.	matr.	# inst.	Matr.	# inst.	Matr.	# inst.	matr.	# inst.	matr.	# inst.	matr.
Escuelas total	121	4623	7	80	1	21	7	260	4	122	7	173	149	5279
Escuelas públicas	36	337	5	28	-	-	2	23	2	23	2	16	47	427
Escuelas privadas	44	1079	3	52	1	21	3	73	1	20	2	35	54	1280
Escuelas D5	41	3207	-	-	-	-	3	164	1	79	3	122	48	3572
Escuelas total	12	722	1	77	1	70	1	86	1	74	1	10	17	1039
Escuelas públicas	2	24	-	-	-	-	-	-	-	-	1	10	3	34
Escuelas D5	10	698	1	77	1	70	1	86	1	74	-	-	14	1005
Escuelas total	1023	53861	698	10500	535	8880	395	11674	287	5470	443	7490	3381	97875
Escuelas públicas	256	1336	367	2245	236	1588	122	847	167	742	237	1164	1385	7922
Escuelas privadas	17	3642	-	-	1	42	4	469	2	537	4	294	28	4984
Escuelas D5	389	7752	286	5124	268	5000	202	4579	91	1722	160	2509	1396	26686
Escuelas D5	361	41131	45	3131	30	2250	67	5779	27	2469	42	3523	572	58283
Escuelas D5*	864	57533	335	8384	300	7383	280	11150	123	4901	212	6493	2115	95844
	-	60	-	8,75	-	7,7	-	11,63	-	5,11	-	6,77	-	100

En este total se excluyen las unidocentes, ya que no se consideran dentro de la muestra. Para efectos de tener completo el panorama general es importante, pero no para el cálculo del número de escuelas en cada celda.

La asignación de la muestra fue proporcional según el número de alumnos en cada grupo, y empleando el redondeo hacia arriba en el caso de los grupos o estratos pequeños. Las 183 instituciones se asignaron en forma proporcional a las seis regiones de MIDEPLAN. En determinados casos, grupos pequeños se fundieron con la finalidad de poder obtener representación de ellos en la muestra, como, por ejemplo, privado D1 y D2 a D5.

Se espera que 183 grupos multiplicados por 35 alumnos (promedio esperado de alumnos por grupo) = 6405 estudiantes en total en la muestra .

Multiplicando los marginales, se determinó el tamaño de muestra en cada celda. La multiplicación se efectuó con la cantidad de alumnos y no con la cantidad de escuelas.

Cuadro 22
Número de escuela en la muestra,
por región, según tipo y tamaño:

	CENTRAL	BRUNCA	H.NORTE	H. ATL.	PA.CTL.	CHOROT.	TOTAL
Privados D1	2	1	1	<u>1</u>	<u>1</u>	<u>1</u>	<u>11</u>
Priv. D2-D5	4	-	-				
Subv. D1.	1	-	-	-	-	1	2
Subv. D2-D5	1	1	1	1	1	-	5
Publ. Urb. marg.	5	-	1	1	1	1	9
Public. D1	30	4	4	6	2	3	49
Publ. D2-D5	64	9	8	13	6	7	107
Total	107	15	15	22	11	13	183

Los listados correspondientes al marco muestral estaban organizados en forma congruente con estos estratos y grupos. Es importante resaltar que los listados contienen solamente la información de la matrícula de primer año. El resto de los niveles no es de interés; por lo tanto, a la base de datos madre se le solicita solamente la matrícula total de primer año, incluso sin dividir por sexo, solamente con otra columna con su acumulación. Además, en el programa se piden los subtotales para cada uno de los grupos o celdas correspondientes al cuadro anterior. La selección se

efectuó en forma sistemática con un arranque aleatorio, en casos de muestras pequeñas, y dos o más para las muestras mayores.

Capítulo 3.

Este capítulo contiene la información correspondiente a todas las pruebas diagnósticas de conocimientos, de los a niveles tercero, sexto y noveno nivel de Educación General Básica, para los diferentes años en que se diseñó la muestra dentro de este proyecto de investigación. Los ajustes o diferencias de año a año son pocas, respecto a la estrategia empleada para organizar el marco muestral y determinar los tamaños de muestra para cada estrato o subgrupo.

2.1. Pruebas diagnósticas de abril y mayo de 1995.

Las pruebas diagnósticas de conocimientos de abril de 1995, implicaba efectuar medición de conocimientos en las cuatro materias básicas. Los niveles educativos por medir eran tercero, sexto y noveno año, de la siguiente forma:

tercero: Estudios Sociales y Ciencias

sexto: Español y Matemáticas

noveno: Estudios Sociales, Ciencias, Español y Matemáticas.

Las estimaciones iniciales alcanzaban cifras de cuarenta mil pruebas, es decir, diez mil por asignatura. El trabajo de campo exigía aplicar ocho pruebas diferentes, según correspondiera al nivel. Se esperaba visitar alrededor de doscientos cincuenta instituciones educativas.

Al visitar una institución de Primaria, se podía aplicar prueba en dos niveles educativos, no así en Secundaria, donde solo un nivel sería medido. Otra decisión que debía ser tomada, a la luz de los recursos, era si cada estudiante de la muestra ejecutaba solo una prueba, dos pruebas o todas las correspondientes a su nivel. Esto significaba que se podían hacer dos pruebas en un mismo día o era necesaria la revisita a la institución. Adicionalmente, otro problema era la selección muestral, si sería aleatoria en cada institución o de conglomerados completos, es decir, de grupo o aula completa.

Las instituciones educativas del país pueden ser agrupadas de diferentes formas, como, por ejemplo, según:

financiamiento: públicas, privadas y subvencionadas

zona: urbana y rural

matrícula: unidocente, dirección 1,2,3,4 o 5

apoyos adicionales: urbano marginal, indígena, informática

ubicación geográfica: dirección regional 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18 y 19

Cuadro # 23	
Clasificación de las instituciones, según matrícula inicial	
Número de alumnos	Tipo de institución
PRIMARIA	
50 o menos	Unidocente
51-180	Dirección 1 (D1)
181-300	Dirección 2 (D2)
301-600	Dirección 3 (D3)
601-900	Dirección 4 (D4)
901- y más	Dirección 5 (D5)
SECUNDARIA	
Menos de 750	Dirección 1 (D1)
751-1500	Dirección 2 (D2)
1501 y más	Dirección 3 (D3)

Se contaba con la experiencia, del año anterior, de haber visitado 180 instituciones, en las cuales se recolectaron treinta mil ochocientas (30800) pruebas en 2 visitas por institución (360 visitas a instituciones). Haciendo una inferencia y proyección bajo el mismo modelo (en el año 94), sabemos que encuestar 6000 estudiantes implica a 701 instituciones, de la siguiente manera:

$$\begin{array}{r} 360 \text{ inst} \quad 30800 \text{ enc.} \\ x \quad \quad \quad 6000 \text{ enc.} \end{array}$$

Con los datos disponibles del Ministerio de Educación Pública, se pudo construir el siguiente cuadro, para medir la viabilidad de las diferentes opciones: encuestar solo una vez a los estudiantes, hacer la revisita o dividir cada aula o institución en grupos para la aplicación de las diferentes pruebas.

	PRIMARIA		SECUND.	TOTAL
	TERCERO	SEXTO	NOVENO	
MATRICULA	87 808	62 709	34 109	184 626
Porcentaje	47,6 %	34,0 %	18,5%	100 %
<i>Posible cantidad de pruebas</i>				
-Español	*****	9 750	5 250	15 000
-Ciencias	10 800	*****	4 200	15 000
-Matemática	*****	9 750	5 250	15 000
-Est. Sociales	10 800	*****	4 200	15 000
Total pruebas	21 600	19 500	18 900	60 000
% de pruebas respecto a la matrícula de su nivel	24,6 %	31,0%	55,4 %	32,5 %

Con un plan de 60 000 pruebas para obtener un número real de 40 000, en promedio, implicaba 1 de cada 3 estudiantes, número considerado imposible de alcanzar con los recursos disponibles .

Otra posibilidad fue partir de 40 000 pruebas y no de 60 000.

Se disponía de 25 días hábiles para el trabajo de campo en instituciones. El personal total se dividió en 5 equipos; cada uno con un coordinador y 3 aplicadores.

Si se visitaban 400 instituciones, debían asignarse 80 instituciones a cada equipo, y si se disponía de 5 semanas, esto correspondía a 20 instituciones por semana, por equipo.

Cálculos generales, basados en las listas del Ministerio de Educación, a partir de las cuales se estructuraría el marco muestral, brindaban los siguientes resultados:

Educación primaria

3544 instituciones en todo el país (según publicación oficial del MEP):

1422 unidocentes

1453 dirección 1

244 dirección 2

217 dirección 3

89 dirección 4

109 dirección 5

3534 total, según datos a partir del marco muestral.

Aquí se da una diferencia de 10 instituciones. No era viable establecer o identificar el error en el sentido de si realmente faltaban 10 instituciones y cuáles eran. Dado el poco tiempo disponible esta diferencia se consideró irrelevante, más bien, indica un alto grado de coincidencia entre el marco muestral y los datos oficiales. La detección del faltante o sobrante implicaba una alta inversión de tiempo para comparar ambos listados, los cuales no estaban ordenados de igual forma. A partir de la información, se desprende que existen cerca de 3000 instituciones pequeñas (82 %) . Las instituciones que se consideran pequeñas en este caso, fueron las unidocentes y las dirección 1 (D1), es decir, instituciones cuya población estudiantil es inferior a 50 alumnos en las unidocentes y a 180 en las dirección 1 (D1). Se espera que estas instituciones, en su mayoría, se espera que estén ubicadas geográficamente en zonas rurales, poco céntricas o alejadas, por lo que el acceso se dificulta. Asimismo, es difícil coordinar las visitas a ellas por falta de teléfono y de otras formas expeditas de rápida comunicación. La cantidad de estudiantes en los niveles de tercero y sexto serán también menores con respecto a las instituciones más grandes. Por lo tanto, el costo de llegar a estas escuelas es más alto y la cantidad de información por es poca. Esto no significa que no sea importante, sino que el acceso es costoso. Soluciones intermedias es una muestra pequeña en estos grupos, asegurando que estén representadas, aunque en menor proporción. Sus resultados se ponderan o bien se tabulan y se presentan separadamente, como un grupo aparte.

En Secundaria, el total de instituciones en al año 95 era de 322: 232 (72%) públicas, 71 privadas y 19 subvencionadas.

Las instituciones, además de estar clasificadas por su tamaño, es decir, por su cantidad de alumnos, se clasifican en otros grupos, según la pertenencia a un proyecto, como urbano-marginal, indígena, rural, informática u otro. Por lo tanto, no solamente interesa la representación por tamaño, sino también por otro tipo de características, las cuales no necesariamente son excluyentes. Esto hace la selección más compleja, por lo que se ha procedido de la siguiente manera: si se necesita representación de urbano -marginales, se selecciona una muestra en forma separada o independiente, y las tabulaciones y presentación de datos se hace, también, en forma separada del resto de los datos o muestra general.

Para determinar la distribución de la muestra por estratos o grupos fue necesaria la producción de los siguientes cuadros. Esto implicó la reelaboración de cuadros de trabajos anteriores y laboriosos, por ser el primer acercamiento a la construcción y por no disponer del material organizado en la forma apropiada.

Cuadro 24
Instituciones unidocentes,
por región según matrícula en tercero y sexto año

REGIÓN	Número de Instituciones	Matrícula	
		Tercero	Sexto
Central	256	1053	645
Brunca	367	1522	904
Huetar Norte	236	1143	600
Huetar Atlántica	122	591	298
Pacífico Central	167	656	405
Chorotega	237	1006	668
Total	1385	5971	3520

Cuadro 25
Instituciones de educación primaria
por región, según tipo de institución:

	Central	Brunca	Huetar Norte	Huetar Atlántico	Pacífico Central	Chorotega	TOTAL
PRIVADOS							
Unidocente							
Número instituciones:	36	5	-	-	2	2	45
Matrícula tercero:	127	21	-	-	18	13	179
Matrícula sexto:	55	6	-	-	4	11	76
PRIVADOS							
Número instituciones:	121	8	1	8	4	7	149
Matrícula tercero:	3685	66	18	213	66	153	4201
matrícula sexto:	3029	41	15	150	43	125	3403
SUBVENCIONADOS							
Número Instituciones:	12	1	1	1	1	1	17
matricula tercero:	694	66	80	47	97	14	998
matrícula sexto:	659	67	69	31	103	11	940
Público D2 a D5.							
Número instituciones:	361	45	30	67	27	42	572
matrícula tercero:	36745	2697	1938	4887	2127	3225	51619
matrícula sexto:	29250	1952	1397	2958	1663	2377	39597

Cuadro 26
Instituciones de educación secundaria,
por región, según tipo:

		Central	Brunca	Huetar Norte	Huetar Atlántico	Pacif. Central	Chor.	TOTAL
Secundaria Privada								
# de Instit.:	D1	56	2	-	5	4	4	71
	D2	1	-	-	-	-	-	1
matrícula:	D1	2342	5	-	113	38	96	2594
	D2	142	-	-	-	-	-	142
Subvencionada Privada								
# de Instit.:	D1	15	-	1	-	-	-	16
	D2	4	-	-	-	-	-	4
matrícula:	D1	1445	-	67	-	-	-	1512
	D2	627	-	-	-	-	-	627
Público								
# de Instit.:	D1	62	31	20	11	18	27	169
	D2	72	4	1	9	3	7	96
Matrícula:	D1	3279	1623	1027	364	1168	1414	8875
	D2	15754	797	281	1755	580	1192	20359
TOTAL:								
# inst.:		210	37	22	25	25	38	357
matrícula:		23589	2425	1375	2232	1786	2702	34109

2.2. Pruebas diagnósticas de conocimientos de noviembre 1995

La muestra para las pruebas diagnósticas de conocimientos de noviembre de 1995, implicaron similar trabajo al realizado a inicio del año, solamente que con el marco muestral obtenido a partir de la base de datos de la matrícula del año 95, suministrado por el MEP. Los estratos y grupos fueron las seis regiones de MIDEPLAN. Las instituciones de educación primaria se dividieron en :

-unidocentes y públicas con menos de 71 estudiantes en la institución

-urbano marginales : pequeñas

grandes

-privadas y subvencionadas

-públicas: pequeñas (Dirección 1 con 71 alumnos o más)

medianas (Dirección 2 y 3)

grandes (Dirección 4 y 5)

En Secundaria, la división se efectuó así:

-privadas y subvencionadas

-públicas: pequeñas (Dirección 1)

medianas (Dirección 2)

grandes (Dirección 3)

Con base en esta clasificación, se generan los siguientes cuadros:

Cuadro 27
Instituciones de Educación Primaria y matrícula
por tipo y nivel, según región

Región	PRIVADAS				SUBVENCIONADAS				TOTAL			
	# esc	Matr. 3	matr. 6	total matr	# esc	matr. 3	matr. 6	total matr.	# esc	matr 3	Matr. 6	Total matr.
CTL	121	3685	3029	6714	12	694	659	1353	133	4379	3688	8067
Brunca	8	66	41	107	1	66	67	133	9	132	108	240
Huetar orte	1	18	15	33	1	80	69	149	2	98	84	182
Huetar tlánt.	8	213	150	363	1	47	31	78	9	260	181	441
efífico	4	66	43	109	1	97	103	200	5	163	146	309
oro- ga	7	153	125	278	1	14	11	25	8	178	125	303
TOTAL	149	4201	3403	7604	17	998	940	1938	166	5199	4343	9542

Cuadro 28

Escuelas unidocentes y públicas, Dirección 1 con 70 o menos estudiantes, por región, según número de escuela y matrícula por nivel:

Región	# de escuelas	Matrícula 3	matrícula 6.	TOTAL
1. Central	399	2466	1631	4097
2. Brunca	516	3100	1825	4925
3.H. Norte	377	2561	1394	3955
4.H. Atlántica	195	1328	710	2038
5.Pacífico Ctl.	212	1115	730	1845
6. Chorotega	331	1964	1312	3276
TOTAL	2030	12534	7602	20136

Cuadro 29

Escuelas públicas. Dirección 1 a Dirección 5, por región, según número de escuelas y matrícula por nivel: (Dirección 1 con 71 o más alumnos)

Región	Dirección 1 (71 o + alumnos)				D2 y D3		D4 y D5		TOTAL	
	# esc.	matr. 3	matr. 6	total	# esc.	total	# esc.	total	# esc.	total matr.
1.Ctl.	246	4870	3229	8099	219	21091	142	44904	607	74094
2.Br.	137	2589	1710	4299	42	3863	3	786	182	8948
3.H.N.	127	2395	1390	3785	26	2387	4	948	157	7120
4.II. Atl.	129	2657	1349	4006	53	4002	14	3843	196	11851
5 P. Ctl.	46	889	543	1432	21	2117	6	1673	73	5222
6. Ch.	65	1185	800	1985	33	3142	9	2460	107	7587
Total	750	14585	9021	23606	394	36602	178	54614	1322	114822

Cuadro 30
Instituciones de Educación Secundaria,
por región, según tipo de financiamiento y nivel:

REGIÓN	PRIVADO		Subvencionado		Sub total		PUBLICO						TOTAL	
	# Esc.	matr. 9.	# Esc.	matr. 9.	# Esc.	matr. 9.	D1		D 2		D 3			
							# Esc.	matr. 9.	# Esc.	Matr. 9.	# Esc.	matr. 9.	# Esc.	Matr. 9.
Central.	57	2484	19	2072	76	4556	62	3279	54	9543	18	6211	134	19033
Brunca.	2	5	-	-	2	5	31	1623	3	423	1	374	35	2420
Huetar Norte	-	-	1	67	1	67	20	1027	-	-	1	218	21	1308
Huetar Antica.	5	113	-	-	5	113	11	384	6	938	3	817	20	2119
Pacífico Central.	4	38	-	-	4	38	18	1168	3	580	-	-	21	1748
Protegea.	4	96	-	-	4	96	27	1414	7	1192	-	-	34	2606
TOTAL	72	2736	29	2139	92	4875	169	8875	73	12676	23	7683	265	29234

Para cada una de las regiones y tipos de división antes mencionada, se generó un programa para producir listas independientes con la información de la matrícula de tercero y sexto o noveno, según correspondiera. Adicionalmente, cada lista, de manera interna, fue ordenada con criterio geográfico por provincia, cantón y distrito. La muestra se seleccionó en forma sistemática, con la finalidad de abarcar las diferentes regiones geográficas. Para efectuar la afijación, se elaboraron matrices con la información de la matrícula de las instituciones en los niveles respectivos (tercero y sexto o noveno).

Para las instituciones unidocentes (junto con las dirección 1 con menos de 71 alumnos) y las urbano marginales, se seleccionó una muestra de 30 instituciones, por un lado, y 8 por otro (4 pequeñas y 4 grandes), con la finalidad de efectuar un análisis separado del resto de la muestra, dadas las características particulares de este tipo de instituciones y su ubicación geográfica.

Para el resto de las instituciones, según la región a que pertenece la muestra, se afijó en forma proporcional, según la matrícula del nivel o niveles en que se efectuaría la aplicación de la prueba.

Se consideraron como datos ciertos los correspondientes a la matrícula de las instituciones, brindada por el MEP.

Se estimó que la pérdida podría ser del 15 %, con base en la experiencia de las pruebas de abril y mayo del 95. Este dato podría estar subestimando la situación real, dado que en el transcurso del año se da deserción y movilidad o traslado.

Con el fin de realizar el cálculo de los efectos de diseño (DE), se supone que:

- las instituciones tienen igual matrícula en tercero y sexto (b), lo cual no necesariamente se cumple.
- desviación estándar de 10 y constante en una escala de 40 puntos, para las pruebas en todas las asignaturas, niveles, regiones y sectores. Este dato se basa en la aplicación de pruebas entre 1986 y 1989.
- $\rho_{oh} = 0,03$.

También se partió de datos iniciales como $N = 190109$

$$A = 3963 \quad \text{promedio} = 48$$

$$a = 400$$

Con la asignación proporcional de la muestra, se obtuvo la siguiente distribución:

Cuadro 31
Número de escuelas en la muestra
por región, según tipo de institución:

	Central	Brunca	H.Norte	H.Atlán.	Pacíf.	Chorot.	Total
PRIMARIA							
Unidocente							30
Urbano-marg.							8
Privadas	18	1	-	1	1	2	23
Subv.							
Públicas:							
Pequeñas	49	27	26	25	9	13	150
Medianas	44	9	5	11	4	7	80
Grandes	29	1	1	3	1	2	37
SECUNDARIA							
Privadas	13	-	-	1	1	1	16
Subv.							
Públicas:							
pequeñas	11	6	4	2	4	6	33
medianas	11	1	-	1	1	1	14
grandes	4		-	1	-	-	6

2.3. Pruebas diagnósticas de conocimientos de noviembre 1996

La prueba diagnóstica de conocimientos de 1996 debió diseñarse partiendo del marco muestral de 1995, dado que debían programarse las giras y coordinar con la sección de transportes de la Universidad, antes del mes de julio. La diferencia de esta muestra respecto a la del año anterior reside en que excluye de la prueba noveno año, es decir, solo se toma en cuenta Primaria en tercero y sexto años, para las cuatro materias básicas. El procedimiento de organización y de selección de la muestra fue similar al de noviembre de 1995 y con el mismo marco muestral.