



**ENCUESTA SOBRE TRAYECTORIAS
EDUCATIVAS Y LABORALES DE LOS
JÓVENES MEXICANOS**

REPORTE FINAL



Instituto Nacional
para la Evaluación
de la Educación

www.inee.edu.mx

Noviembre 2010

CONTENIDO

	<i>Página</i>
1. Antecedentes	3
2. Síntesis del diseño y selección de la muestra	
2.1 Diseño inicial de la muestra	3
2.2 Ajustes al diseño inicial	5
2.3 Probabilidades de selección	6
2.4 Factores de Ajuste Demográfico (estructuras de sexo y edad)	8
3. Logística de campo.	
3.1 Selección de viviendas	10
3.2 Selección de informantes	11
4. Cuestionario y proceso de capacitación	11
5. Resultados del trabajo de campo	13
6. Codificación de preguntas abiertas	17
7. Crítica, Captura y validación de los cuestionarios	18
Anexo 1. Muestra seleccionada	
Anexo 2. Formato Listado de Vivienda	
Anexo 3. Cuestionario	
Anexo 4. Manual del Entrevistador	
Anexo 5. Boletín de campo	
Anexo 6. Listado de Códigos	
Anexo 7. Validaciones	
Anexo 8. Notas sobre la captura	

1. ANTECEDENTES.

El Instituto Nacional para la Evaluación de la Educación (INEE) realizó la **Encuesta de Trayectorias Educativas y Laborales de los Jóvenes Mexicanos 2010**, para lo cual invitó a **Berumen y Asociados, S.A de C.V.** (en adelante **Berumen**) a colaborar en algunas actividades asociadas a la encuesta objeto de este informe.

El **objetivo general** de la encuesta es conocer cuáles han sido las experiencias educativas y laborales de los jóvenes mexicanos. La **población objetivo** son los jóvenes residentes del Área Metropolitana de la Ciudad de México (**AMCM**) que tengan entre **18 y 29** años de edad.

De acuerdo a lo convenido con el **INEE**, durante todo el desarrollo del estudio participó un grupo de asesores académicos del Colegio de México (**COLMEX**), quienes mantuvieron facultades de supervisión y seguimiento del contrato mediante la observación directa y presencia en diferentes partes del proceso: el diseño de la muestra, la capacitación, el trabajo de campo, la codificación, el procesamiento y la entrega de la Base de Datos.

2. SINTESIS DEL DISEÑO Y SELECCIÓN DE LA MUESTRA.

2.1) Diseño inicial de la muestra.

A solicitud del **INEE** se seleccionó una muestra probabilística del Área Metropolitana de la Ciudad de México (**AMCM**) que al final del operativo diera como resultado alrededor de **2,900** entrevistas completas, con los informantes debidamente clasificados en cuatro categorías de niveles socioeconómicos (**NSE**), con un nivel de precisión del orden de **+2%** para las estimaciones globales de **parámetros prioritarios**, y alrededor de entre **+3 a 6%** en cada categoría de **NSE** según el número de casos que resultaran en muestra en cada nivel, todas con un nivel de confianza del **95%** en todos los casos.

Con estos parámetros se diseñó una muestra en cuatro etapas apegada a los siguientes lineamientos generales:

- 1) En la primera etapa se seleccionaron **280** Áreas Geoestadísticas Básicas (**AGEB**) con probabilidad proporcional a una medida de tamaño (**PPT**) de acuerdo al Censo de Población y Vivienda del **2005** (ya que aún no están disponibles las cifras del Censo **2010**). El marco muestral lo integran el total de **AGEB** de las delegaciones y municipios que conforman el Área Metropolitana de la Ciudad de México. La muestra seleccionada a nivel **AGEB** se presenta en el **ANEXO 1**.

En el cuadro **1** se muestra la distribución de las **280** **AGEB** en el **AMCM** así como el número de viviendas a seleccionar para entrevista según el diseño inicial.

Cuadro 1. Distribución de la muestra

Núm.	Edo	Clave	Municipio / Delegación	AGEB en muestra	Viviendas a seleccionar
1	09	002	Azcapotzalco	7	105
2	09	003	Coyoacán	11	165
3	09	004	Cuajimalpa	3	45
4	09	005	Gustavo A. Madero	19	285
5	09	006	Iztacalco	6	90
6	09	007	Iztapalapa	27	405
7	09	008	Magdalena Contreras	3	45
8	09	009	Milpa Alta	1	15
9	09	010	Álvaro Obregón	11	165
10	09	011	Tláhuac	5	75
11	09	012	Tlalpan	9	135
12	09	013	Xochimilco	6	90
13	09	014	Benito Juárez	8	120
14	09	015	Cuahutémoc	10	150
15	09	016	Miguel Hidalgo	7	105
16	09	017	Venustiano Carranza	7	105
Subtotal Distrito Federal				140	2100
17	15	013	Atizapan de Zaragoza	9	135
18	15	020	Coacalco de Berriozabal	6	90
19	15	024	Cuautitlán	2	30
20	15	025	Chalco	2	30

Núm.	Edo	Clave	Municipio / Delegación	AGEB en muestra	Viviendas a seleccionar
21	15	029	Chicoloapan	3	45
22	15	030	Chinconcuac	1	15
23	15	031	Chimalhuacán	8	120
24	15	033	Ecatepec de Morelos	29	435
25	15	039	Ixtapaluca	5	75
26	15	053	Melchor Ocampo	1	15
27	15	057	Naucalpan de Juárez	14	210
28	15	058	Nezahualcóyotl	20	300
29	15	059	Nextlalpan	1	15
30	15	060	Nicolás Romero	4	60
31	15	070	La Paz	4	60
32	15	091	Teoloyucan	1	15
33	15	095	Tepotztlán	1	15
34	15	099	Texcoco	1	15
35	15	104	Tlanepantla de Baz	13	195
36	15	108	Tultepec	1	15
37	15	120	Zumpango	1	15
38	15	121	Cuautitlán Izcalli	8	120
39	15	122	Valle de Chalco Solidaridad	5	75
Subtotal Estado de México				140	2100
Gran total				280	4200

- 2) En la segunda etapa se seleccionaron de manera aleatoria **dos** manzanas en cada una de las **AGEB** en muestra.
- 3) En la tercera etapa, se seleccionó una muestra aleatoria de **7** viviendas particulares ocupadas en una de las manzanas y **8** en la otra manzana, para un total de alrededor de **4,100** viviendas seleccionadas para el **AMCM**.
- 4) Se seleccionó de manera aleatoria un solo hogar en aquellas viviendas en muestra con más de uno.
- 5) En la última etapa de muestreo fue seleccionado un miembro del hogar entre **18** y **29** años mediante el método de Kish.

El diseño probabilístico no acepta sustituciones, y por experiencias similares se esperaba produjera una muestra global estrictamente probabilística de **alrededor de 4,100** viviendas particulares, que con una tasa estimada de no-respuesta del orden de un poco menos del **30%**, se esperaba produjeran alrededor de **2,900** entrevistas completas deseadas en el **AMCM**.



Por los niveles de inseguridad actuales en la Ciudad de México se estimó una tasa de no-respuesta del orden del **25-30%** para zonas donde predominen los **NSE C, D+ y D/E**, y del orden de **40-50%** donde predominen los niveles **ABC+**.

2.2) Ajustes al diseño inicial.

Con el objetivo de observar la aplicación del cuestionario en campo, del **25** al **27** de septiembre de **2010** el personal capacitado realizó la primera etapa de levantamiento en **AGEB** del Distrito Federal.

El principal hallazgo en estos tres primeros días del operativo fue identificar que la selección de segmentos de **7** y **8** viviendas por manzanas **no** era suficiente para lograr las entrevistas esperadas ya que en más del **50%** de las viviendas **no** existía población objetivo. Por lo anterior se hizo un primer ajuste a la logística operativa, que consistió en hacer un barrido de toda la manzana seleccionada en lugar de tocar sólo un segmento de **7** y **8** viviendas.

Por lo anterior, a partir del martes **28** de septiembre se ajustó la metodología de campo pidiendo a los encuestadores que hicieran un “**barrido**” de las **2** manzanas seleccionadas por **AGEB**, esto es, que tocaran cada una de las viviendas para identificar aquellas con población objetivo y proceder a incluirlas en la muestra.

Con cifras de campo al día lunes **4** de octubre se apreció que para lograr cubrir la meta deseada (alrededor de **2,900** entrevistas) se requería tocar más viviendas, por lo que se realizó un segundo ajuste a la logística operativa. En los **AGEB** que se empezaron a trabajar a partir de esta fecha se seleccionaron **2** manzanas mas, por lo que en total serían **4** manzanas por **AGEB**. Con lo anterior se estimó que se tocaría alrededor de **30,000** viviendas para lograr la meta de alrededor de **2,900** entrevistas.

Al final del trabajo de campo, el número de manzanas seleccionadas por **AGEB** quedaron distribuidas según se muestra en el siguiente cuadro.

Cuadro 2. Manzanas seleccionadas por AGEB

Entidad Federativa	AGEB con 2 manzanas	AGEB con 4 manzanas	AGEB seleccionados
Distrito Federal	85	55	140
Estado de México	0	140	140
Total	85	195	280

2.3) Probabilidades de Selección.

De los **280 AGEB** seleccionadas, **140** corresponden al Distrito Federal y **140** a municipios del Estado de México que son parte del **AMCM**. Para definir las probabilidades de selección, y con ello determinar los factores de expansión correspondientes sean:

D_h : el estrato **h** (delegación o municipio) dentro del dominio **D** (DF o municipios conurbados del Estado de México).

S_{Dh} : el número de **AGEB** seleccionadas en el estrato **h** del dominio **D**; en total **140 AGEB** para el Distrito Federal y **140** para los municipios conurbados del Estado de México.

P_{Dhi} : la probabilidad de selección de la **i**-ésima **AGEB** del estrato **h** para el dominio **D**.

P_{Dhij} : la probabilidad de selección de la **j**-ésima manzana dentro de la **i**-ésima **AGEB** del estrato **h** para el dominio **D**.

N_{Dhi} : el total de viviendas dentro de la **i**-ésima **AGEB** del estrato **h** del dominio **D**.

N_{Dhij} : el total viviendas dentro de la **j**-ésima manzana de la **i**-ésima **AGEB** del estrato **h** del dominio **D**.

M_{Dhi} : el total manzanas dentro de la **i**-ésima **AGEB** del estrato **h** del dominio **D**.

N_{Dh} : el total de viviendas del estrato **h** en el dominio **D**.

N_D : el total de viviendas en el dominio **D**.

La probabilidad de seleccionar el **AGEB i** del estrato **h** en el dominio **D** está entonces dada por:

$$P_{Dhi} = \frac{S_{Dh} * N_{Dhi}}{\sum_{i=1} N_{Dhi}} = \frac{140 * N_{Dhi}}{N_{Dh}}$$

Y la probabilidad de seleccionar la manzana **j** del **AGEB i** del estrato **h** en el dominio **D** está dada por:

$$P_{Dhij} = \frac{k}{M_{Dhi}}, \quad \text{donde } k \text{ es } 2 \text{ o } 4 \text{ según el caso (número de manzanas seleccionadas dentro del AGEB)}$$

Por lo tanto, la probabilidad final de selección para cualquier hogar “I” dentro de la manzana j de la AGEB i del estrato h del dominio D está dada por:

$$P_{Dhijl} = P_{Dhi} \circ P_{Dhij} \circ 1$$

$$= \frac{140 N_{Dhi}}{N_{Dh}} * \frac{k}{M_{Dhi}} * 1$$

La variable **PONDH** (el inverso de la probabilidad anterior) que se incluye en el archivo de datos, es el valor que se debe utilizar para el cálculo de estimadores de razón referentes a hogares.

Al interior de los hogares los integrantes de **18 a 29** años, fueron seleccionados con igual probabilidad de selección dependiendo del número de ellos en cada hogar, por lo que es necesario usar el ponderador correspondiente a cada individuo-hogar en muestra.

La probabilidad de que un integrante del hogar entre **18 y 29** años fuera incluido en la muestra es:

$$P_{Dhijl} = \frac{1}{Y_{Dhijl}}$$

donde:

Y_{Dhijl} \equiv número de integrantes entre **18 a 29** años dentro del hogar I de la manzana j del AGEB i del estrato h del dominio D

La variable **PONDI** (el inverso de la probabilidad anterior) que se incluye en el archivo de datos, es el valor que se debe utilizar para el cálculo de estimadores de razón referentes a Individuos en el hogar, ésta incluye tanto la probabilidad a nivel hogar como la probabilidad individual de hogar.

Las estimaciones de la encuesta no son el resultado directo de las entrevistas levantadas. Son el cómputo para la estimación de cada parámetro e indicador de interés que toma en cuenta:

- El factor de expansión a nivel hogar de acuerdo al diseño y selección de viviendas particulares.
- El factor de expansión correspondiente al miembro del hogar particular seleccionado.
- Factores de corrección por no respuesta a nivel hogar.
- Factores de corrección por no respuesta a nivel individual.
- Factores de ajuste demográfico (estructuras de sexo y edad).



2.4) Factores de Ajuste Demográfico (Estructuras de sexo y edad).

Se verificaron en fuentes exógenas a la encuesta las estructuras correspondientes según subgrupos de sexo y edad para con ello determinar los factores de ajuste demográfico. De acuerdo a lo solicitado por el **INEE**, se tomaron los resultados del **INEGI** de la Encuesta Nacional de Ocupación y Empleo (**ENOE**), de la Encuesta Nacional de Ingresos y Gastos de los Hogares (**ENIGH 2008**) así como el Censo de Población y Vivienda **2005**, los que se muestran en los siguientes cuadros:

Cuadro 3.1
Estructura de la ENOE según grupo de edad y sexo

Rango de edad	Sexo		Total
	Hombres	Mujeres	
18 - 20 años	29.5%	26.2%	27.8%
21 - 23 años	25.4%	25.6%	25.5%
24 - 26 años	22.5%	24.2%	23.4%
27 - 29 años	22.6%	23.9%	23.3%
Total	100%	100%	100%

Cuadro 3.2 Estructura del Censo de Población y Vivienda 2005 según grupo de edad y sexo

Rango de edad	Sexo		Total
	Hombres	Mujeres	
18 - 20 años	28.9%	25.3%	27.1%
21 - 23 años	24.6%	23.3%	24.0%
24 - 26 años	22.7%	24.8%	23.8%
27 - 29 años	23.8%	27.5%	25.7%
Total	100%	100%	100%

Cuadro 3.3
Estructura de la ENIGH según grupo de edad y sexo

Rango de edad	Sexo		Total
	Hombres	Mujeres	
18 - 20 años	30.3%	25.9%	28.1%
21 - 23 años	25.4%	22.9%	24.2%
24 - 26 años	21.4%	25.0%	23.2%
27 - 29 años	23.0%	26.1%	24.6%
Total	100%	100%	100%

El promedio que resulta de las tres fuentes exógenas se muestra a continuación:

Cuadro 3.4 Estructura promedio de datos exógenos según grupo de edad y sexo

Rango de edad	Sexo		Total
	Hombres	Mujeres	
18 - 20 años	29.6%	25.8%	27.7%
21 - 23 años	25.1%	23.9%	24.5%
24 - 26 años	22.2%	24.7%	23.4%
27 - 29 años	23.1%	25.8%	24.5%
Total	100%	100%	100%

La que resulta de la encuesta con datos ya ponderados es la siguiente:

Cuadro 3.5 Estructuras según la Encuesta de Trayectorias Educativas y Laborales DF 2010

Rango de edad	Sexo		Total
	Hombres	Mujeres	
18 - 20 años	39.2%	32.4%	35.8%
21 - 23 años	25.9%	26.1%	26.0%
24 - 26 años	19.3%	24.9%	22.1%
27 - 29 años	15.6%	16.7%	16.1%
Total	100%	100%	100%

Por último, los factores de ajuste que se propone son los siguientes:

Cuadro 3.6 Factores de Ajuste

Rango de edad	Factores de ajuste	
	Hombre	Mujer
18 - 20 años	0.754141	0.797082
21 - 23 años	0.968754	0.918886
24 - 26 años	1.152008	0.990488
27 - 29 años	1.484361	1.551370

3. LOGÍSTICA DE CAMPO

3.1) Selección de Viviendas.

Siguiendo el diseño y selección de la muestra, en cada Área Geoestadística Básica (**AGEB**) se seleccionaron dos o cuatro manzanas en gabinete. Durante el operativo de campo, las tareas básicas que se realizaron en cada una de las manzanas seleccionadas fueron las siguientes:

- Barrido censal de las viviendas particulares ocupadas dentro de la manzana para aplicar un filtro.
- Aplicación del filtro para detectar viviendas particulares con población objetivo.
- Inclusión en la muestra de las viviendas con hogares con población objetivo.

Para fines del estudio se tomaron en cuenta los siguientes conceptos:

Vivienda particular. Espacio físico con entrada independiente destinada a servir de alojamiento a uno o más hogares, puede estar habitada o deshabitada en el momento de realizar la visita.

Hogar. Conjunto de personas que hacen vida en común dentro de una vivienda, unidas o no por parentesco, que comparten los gastos de manutención y preparan los alimentos en la misma cocina. Comparten un ingreso destinado a cubrir un sustento.

Berumen diseñó un formato de control de campo para llevar un registro de las viviendas seleccionadas en cada manzana y su resultado. En el **Anexo 2** se presenta el “Listado de Vivienda”, formato que permite llevar un control de los resultados de cada manzana seleccionada en la muestra por vivienda y código de resultado.

El “Listado de Vivienda” es un formato en el que se reportan las viviendas particulares (ocupadas) de la manzana y registra todos aquellos datos que permiten su correcta identificación en campo. Para el caso particular de esta encuesta, el formato permitió identificar a las viviendas de interés, esto es, todas aquellas viviendas que tuvieran población objetivo (al menos un hombre o mujer de **18 a 29** años).

La aplicación de la encuesta se apegó a los siguientes criterios operativos:

1. Se tocó en cada una de las viviendas de las manzanas seleccionadas para identificar la existencia de informantes adecuados, hombres y mujeres de **18 a 29** años de edad.
2. No se realizaron sustituciones de manzanas en caso de no respuesta.
3. Se realizaron hasta cuatro visitas a los hogares seleccionados en horarios y días distintos, para lograr la entrevista, registrando cada una de sus visitas en la carátula del cuestionario.

3.2) Selección de informantes.

El informante adecuado para responder el cuestionario es cualquier miembro del hogar entre **18** y **29** años de edad seleccionado mediante el “**método de Kish**”. Al elegir al miembro del hogar a encuestar, se tomó en cuenta lo siguiente:

- a) El informante adecuado era un miembro del hogar seleccionado en muestra.
- b) No eran informantes adecuados las personas fuera del rango de edad (**18 a 29** años), visitantes de la familia, empleados domésticos o personas impedidas física o mentalmente para proporcionar la entrevista.
- c) No había sustitución de informantes seleccionados. Sólo se removían de la lista de elegibles los informantes con alguna discapacidad mental para ser entrevistados, por ejemplo Síndrome de Down.

4. CUESTIONARIO Y PROCESO DE CAPACITACIÓN.

El **INEE** proporcionó a **Berumen** el cuestionario que se aplicó durante el operativo de campo. Para asegurar un esquema de selección estrictamente probabilístico así como una distribución natural apegada a la estructura de edad y sexo de la población objetivo, **Berumen** propuso que en el cuestionario final se integraran las adecuaciones para seleccionar al miembro del hogar de **18 a 29** años de edad mediante el “método de Kish”.

Asimismo, de acuerdo a lo convenido con el **INEE**, en todos los cuestionarios se incluyó un módulo que aplica las preguntas necesarias para clasificar a los informantes en cada uno de los cuatro niveles socioeconómicos (**ABC+, C, D+ y DE**) de acuerdo a la Regla de la Asociación Mexicana de Agencias de Investigación de Mercados y Opinión Pública, A.C. (**AMAI**).

La clasificación de **NSE** de cada uno de los hogares de los jóvenes encuestados se incluye en la base de datos en el campo denominado “NIVEL”.

Además de los datos de identificación (control) del cuestionario, y de la sección de selección del informante, el cuestionario final aplicado en campo recopiló información sobre los siguientes temas:

- Sección I. Características sociodemográficas.
- Sección II. Origen Social.
- Sección III. Historia de Progresión Escolar.
- Sección IV. Historia Laboral.
- Sección V. Historia Familiar.

- Sección **VI**. Historia Residencial.
- Sección **VII**. Cierre de trayectoria educativo
- Sección **VIII**. Trayectoria educativa (contexto institucional).
- Sección **IX**. Trabajo actual o último trabajo.
- Identificación del nivel socioeconómico del hogar.

En el **Anexo 3** de este reporte se presenta el cuestionario aplicado en campo que incluye los filtros para determinar el nivel socioeconómico y la selección del informante adecuado. [El cuestionario tuvo un tiempo promedio de aplicación de 31 minutos.](#)

El personal de campo que participó en el levantamiento de la encuesta fue entrenado durante una capacitación en la que se abarcaron los siguientes temas:

- a) Objetivos y alcances del estudio.
- b) Logística operativa: selección de viviendas y selección del informante (método de Kish)
- c) Instrucciones de llenado del cuestionario y conceptos básicos.

Dada la complejidad del cuestionario en la sección de historias educativas y laborales, la capacitación de gabinete al personal de campo se realizó en dos días: **23 y 24** de septiembre de **2010**. Los asesores académicos del **COLMEX** instruyeron acerca de los alcances e importancia del proyecto y participaron durante todo el proceso aclarando dudas a los participantes. Las instrucciones de llenado del cuestionario, la logística operativa y la administración de la encuesta estuvieron a cargo de **Berumen**. Para complementar y apoyar el curso de capacitación se elaboró el “*Manual del Encuestador*” (ver **Anexo 4**).

Posterior a los días de capacitación se llevaron a cabo dos reuniones de retroalimentación con el personal de campo:

- La primera el día **28** de septiembre. Las entrevistas realizadas durante los tres primeros días del operativo de campo (**25 a 27** de septiembre) sirvieron como capacitación “In situ” y permitieron resolver las dudas del personal de campo durante la aplicación del cuestionario.
- La segunda el día **5** de octubre. Como resultado de una revisión en gabinete de los primeros cuestionarios aplicados en campo se identificaron aquellas preguntas que era necesario reforzar a los encuestadores durante el llenado de los cuestionarios. Adicional a la reunión de retroalimentación de esta fecha se elaboró un boletín de campo para todo el personal [de campo](#) (ver **anexo 5**).

5. RESULTADOS DEL TRABAJO DE CAMPO.

El trabajo de campo se realizó del **25** de septiembre al **2** de Noviembre de **2010**. El personal trabajó en equipos de **5** encuestadores con un supervisor, todos bajo las instrucciones de un coordinador de campo. Todo el personal de campo estuvo debidamente identificado mediante un gafete que contenía los datos personales y fotografía reciente. En total se contó con la participación de **6** equipos.

El supervisor era el enlace entre el personal de su equipo y el coordinador. De forma general, el supervisor tuvo a su cargo la preparación y organización del trabajo de campo, así como la observación, revisión y verificación de las actividades de cada uno de los entrevistadores que integraban su equipo.

Durante el trabajo de campo se realizaron las siguientes actividades de supervisión:

- a) Supervisión directa. A través de acompañamientos a cada uno de los encuestadores, el supervisor verificó la adecuada localización de las viviendas en muestra, una correcta selección del informante, así como una adecuada presentación y aplicación del cuestionario. En caso de detectar alguna falla, el supervisor reforzaba instrucciones al encuestador y si era necesario regresaba al hogar encuestado a rescatar información.
- b) Supervisión posterior. Durante todo el trabajo de campo, el supervisor seleccionó al azar cuestionarios de cada uno de los encuestadores para hacer visitas a las viviendas seleccionadas. Con esto, verificó que los encuestadores visitaron las viviendas y entrevistaron al informante adecuado, a través de la aplicación de algunas preguntas.

Si bien, el compromiso inicial con el **INEE** era supervisar al menos el **10%** del trabajo de campo, como se aprecia en la siguiente tabla, en campo se supervisó el **47.9%** del total de las entrevistas efectivas, cifra que incluye tanto la supervisión directa (al momento de la entrevista) como la supervisión posterior.

Cuadro 4. Porcentaje de entrevistas supervisadas

Tipo de supervisión	Cuestionarios	Porcentaje
Directa	551	18.9%
Posterior	849	29.1%
Total supervisado	1,400	47.9%
Total de cuestionarios	2,920	

Durante el barrido de **AGEB** se aplicó un primer filtro utilizando el formato “Listado de Vivienda”, para identificar aquellas viviendas de interés para la encuesta. Como se muestra en el siguiente cuadro, al final del operativo de campo se tocaron **32,379** viviendas de las cuales en **23,869 (73.7%)** no había ninguna persona entre **18** y **29** años (informantes elegibles), esto es,



no eran parte de la población objetivo. En el **14.3%** de las viviendas se pudo corroborar que existiera al menos un miembro del hogar de **18 a 29 años**

Cuadro 5. Clasificación de Hogares

Clasificación de Hogares	Número de Hogares	Porcentaje
Hogares de interés	4,650	14.3%
Viviendas con posible elegible <u>a/</u>	2,225	6.9%
No hay elegible <u>b/</u>	23,865	73.7%
No es vivienda (oficinas, negocios)	247	0.8%
Viviendas sin acceso <u>c/</u>	865	2.7%
Desocupada	538	1.6%
Total	32,390	100%

a/ Son viviendas ocupadas donde no se pudo determinar si correspondían a hogares de interés ya sea por ausencia de sus ocupantes o por negativa de los informantes a proporcionar información.

b/ Los residentes del hogar o algún vecino informaron que en los vivienda no existía al menos un hombre o mujer entre 16 y 29 años de edad.

c/ Se trata de viviendas con acceso restringido, ya sea porque hay vigilancia o son edificios o calles cerradas.

Una vez que en campo se identificaron las viviendas con población objetivo, los encuestadores realizaron hasta cuatro visitas para lograr la entrevista. Uno de los mecanismos utilizados en el trabajo de campo para controlar la muestra en términos de entrevistas completas y “no respuesta” es registrar el código de resultado. Al final de cada visita a los hogares, los encuestadores registraban el resultado utilizando los siguientes códigos

- 1. Entrevista completa.** Se circuló cuando se obtuvo la información completa del cuestionario.
- 2. Entrevista incompleta.** Este código se asignó cuando se obtuvo información parcial del cuestionario debido a que él o la informante interrumpieron la entrevista por alguna causa.
- 3. Ausente temporal.** Se circuló cuando el elegible del hogar se encontraba fuera (vacaciones, trabajo, estudio, enfermedad, etc.) y no se localizó durante el tiempo que duró el trabajo de campo en la localidad.
- 4. Ausente en el momento de la visita.** Este código se registró cuando el elegible se encontraba fuera de la vivienda por ir a la escuela, al trabajo, etc. y no pudo ser localizado por el encuestador durante las visitas a la vivienda.
- 5. Cita / Aplazada.** Se circuló este código cuando por alguna causa (enfermedad, tiempo, accidente, etc.) el entrevistado no podía contestar, pero proporcionó alguna fecha y hora para poder retornar a aplicar el cuestionario correspondiente.



- 6. Se negó a dar información.** El código aplicó cuando él o la informante se negó a proporcionar información, a pesar de las amplias explicaciones sobre el objetivo de la encuesta.
- 7. Nadie en casa.** Este código consideró los casos en que al llegar a la vivienda no se encontró a ningún informante.
- 8. Informante inadecuado.** Se asignó este código cuando no se encontró al elegible o a personas que no supieron dar información por ser menores de **18** años o bien no son residentes de la vivienda.

De las **4,650** viviendas con población de interés, se lograron **2,920** entrevistas completas, registrando una tasa de No Respuesta a nivel de hogar de **37.2%**. En el cuadro **6** se presenta la distribución de las entrevistas levantadas según código de resultado

Cuadro 6. Distribución de entrevistas según código de resultado

Código de Resultado	Número de Hogares	Porcentaje
1. Entrevista completa	2,920	62.8%
2. Entrevista incompleta	6	0.1%
3. Ausente Temporal	21	0.5%
4. Ausente en el momento de la visita	838	18.0%
6. Se negó a dar información	794	17.1%
7. Nadie en casa	62	1.3%
8. Informante inadecuado	9	0.2%
Total	4,650	100%

La encuesta estimó que en los hogares de los encuestados existían en promedio **1.5** jóvenes elegibles. Como se muestra en el siguiente cuadro, en el **62.4%** de los hogares sólo residía un joven de la edad de interés (**18 a 29** años):

Cuadro 7. Distribución de entrevistas completas según número de elegibles en el hogar

Total de elegibles en los hogares encuestados	Entidad					
	D.F.		Estado de México		Total AMCM	
	Entrevistas	Porcentaje	Entrevistas	Porcentaje	Entrevistas	Porcentaje
1	729	60.3%	1,093	63.9%	1,822	62.4%
2	352	29.1%	442	25.8%	794	27.2%
3	97	8.0%	131	7.7%	228	7.8%
4 a 6	31	2.6%	45	2.6%	76	2.6%
Total	1,209	100%	1,711	100%	2,920	100%
Elegibles promedio	1.54%		1.50%		1.51%	



Debido a que no se podía sustituir a los informantes seleccionados en los hogares, el personal de campo realizó hasta cuatro visitas a los hogares de los informantes seleccionados para lograr la entrevista completa.

Durante el levantamiento de la encuesta **Berumen** entregó reportes semanales al **INEE** para informar el avance del operativo así como los ajustes que se realizaron a la logística operativa para lograr la muestra esperada de **2,900** entrevistas completas.

En términos generales, durante el operativo de campo se registraron las incidencias que se describen a continuación:

1. Una vez que se habían identificado viviendas con población objetivo, se registró una negativa a proporcionar información de los elegibles (sexo y edad exactos).
2. Dificultad para localizar a los jóvenes de mayor edad (**22** años y más) porque muchos de ellos estudian y trabajan. Se reportaba que salían muy temprano y llegaban muy noche (incluso después de las **10** de la noche), que no tenían hora de regreso o que sólo se les localizaba los fines de semana.
3. Una vez seleccionado el elegible en el hogar, los jóvenes de mayor edad y los hombres (en comparación a las mujeres) registraron mayores tasas de no respuesta, ya sea por rechazo a la entrevista o por la dificultad para localizarlos.
4. Rechazo de los hogares a participar en la encuesta, básicamente por lo siguiente:
 - a) Los habitantes de las viviendas en muestra manifestaban que no les interesaba.
 - b) En algunas áreas de alto nivel no se pudo acceder ya sea porque existía vigilancia o por ser áreas cerradas.
 - c) Desconfianza por los problemas de inseguridad del **DF**.
5. La inseguridad en algunos **AGEBs** y los horarios irregulares de los informantes para sus actividades (trabajo, estudio) fueron factores que afectaban para cubrir las citas nocturnas que permitían localizar a las personas que sólo se encuentran de noche.

Al concluir el operativo de campo, se identificaron **53** manzanas completas en las que no se aplicó la encuesta por alguno de los siguientes motivos:

Manzanas	Motivo
14	No hubo acceso en toda la manzana por ser zonas de nivel socioeconómico alto AB/C+ con restricciones para su acceso.
28	La manzana seleccionada tenía uso comercial y/o de servicios: oficinas, negocios, escuelas, clínicas, etc.
10	No había viviendas, toda la manzana era lote baldío o parque.
1	Zona de alto riesgo para el personal de campo.

6. CODIFICACIÓN DE PREGUNTAS ABIERTAS.

El cuestionario incluyó preguntas sobre ocupación para lo cual se acordó con el personal del INEE y los asesores académicos del COLMEX codificar la información a nivel de subgrupo (tres dígitos) utilizando la **Clasificación Mexicana de Ocupaciones (CMO) 2009** del INEGI. Las preguntas sobre empleo que fueron codificadas son:

- **2.7)** *¿Cuál era el nombre del oficio, puesto o cargo de tu -MENCIONE PERSONA REGISTRADA EN PREGUNTA 2.5- cuando tú tenías 15 años?*
- **4.6)** *¿Cuál era el nombre de ese oficio, puesto o cargo?*

Es importante mencionar que las preguntas que indagaban sobre las tareas desempeñadas no fueron codificadas ya que se utilizaron como información de apoyo para codificar la ocupación, se trata de las preguntas:

- **2.6)** *En ese trabajo, ¿Cuáles eran las tareas o funciones principales que desempeñaba tu -MENCIONE PERSONA DE PREGUNTA 2.5- cuando tú tenías 15 años?*
- **4.5)** *¿Cuáles eran las principales tareas o funciones que desempeñabas cuando comenzaste en ese trabajo?*

En el caso de los trabajos de los informantes, también se utilizaron como apoyo para clasificar la ocupación, las preguntas:

- **4.7)** *(Contándote tú, ¿cuánta gente trabajaba aproximadamente en la empresa, negocio o institución en la que tenías ese trabajo?)*
- **4.8)** *(¿En ese trabajo eras...?)*

El cuestionario registró también información sobre localidades, municipios y entidades federativas, la cual fue codificada utilizando la nomenclatura del **Conteo de Población y Vivienda 2005** del INEGI. Dichas preguntas son:

- **6.1)** *(¿En qué localidad -poblado o ciudad- vivías cuando cumpliste cinco años de edad?)*
- **6.2)** *(¿En qué municipio –delegación- estaba esa localidad?)*
- **6.3)** *(¿En qué estado de la república o país estaba esa localidad?)*
- **6.5)** *(¿Podrías decirme todas las localidades -pobladitos o ciudades- donde viviste después de -MENCIONE LOCALIDAD DE RESIDENCIA REGISTRADO EN 6.1- así como la edad que tenías cuando comenzaste a vivir en ese lugar?)*

Otro tipo de preguntas abiertas fueron las que indagaron sobre la carrera en que se inscribió el informante, como las preguntas:

- **8.39)** *(¿Cuál era el nombre de la carrera en la que te inscribiste por primera vez?)*
- **8.41)** *(¿Cuál era el nombre de la carrera en la que te inscribiste por última vez?)*

Para codificar estas preguntas se consultó en Internet la página de la **UNAM** para clasificar las carreras por áreas.



Por último, aunque la pregunta **8.45** (*¿Cuál era el nombre de la última escuela, instituto o universidad en la que cursaste la educación superior?*) no es una pregunta abierta como tal, si aparecieron muchas universidades en la categoría de otros, por lo que estas universidades se codificaron distinguiendo universidades públicas de privadas. Para lograr esta clasificación y validar la información obtenida, se consultó el Directorio Nacional de Instituciones de Educación Superior de la Asociación Nacional de Universidades e Instituciones de Educación Superior (ANUIES).

El listado completo de códigos utilizado se encuentra en el **Anexo 6** de este reporte.

7. CRITICA, CAPTURA Y VALIDACION DE LOS CUESTIONARIOS.

Dadas las características del estudio, para la captura y validación de la encuesta, **Berumen** utilizó el software **CSPro** (Census and Survey Processing System). Este software garantiza que las aplicaciones de captura de datos permitan integrar la información recopilada en campo y hacer una validación individual, sistemática y confiable de los datos. A su vez, posibilita detectar y evitar errores de campo o de diseño.

El **CSPro** permite restringir valores de entrada en cada variable, definir el valor de la “No Respuesta” como un valor especial, desplegar mensajes en el momento de la captura de acuerdo a condiciones que adviertan al capturista de un posible error de captura o inconsistencia, además de permitir definir saltos de un campo a otro de acuerdo a la secuencia lógica del cuestionario.

La validación individual de los cuestionarios a través de las aplicaciones de captura de **CSPro** garantiza la consistencia de los datos recopilados durante el levantamiento de campo, lo que permite obtener una base de datos limpia de errores, inconsistencias y registros duplicados.

En este caso también se ejecutó una validación especial que verificó la consistencia de la trayectoria educativa del entrevistado, los casos con discrepancias se revisaran uno por uno.

Los criterios de validación del cuestionario que fueron integrados al programa de captura, así como los Listados de Códigos utilizados para las preguntas abiertas se presentan los **Anexos 6** y **7**.

Una vez terminada la fase de captura y validación de información se convirtió la base maestra en las bases de datos requeridas por el **INEE** (trayectoria educativa, trayectoria laboral, historial familiar etc.). Estas son las bases que se entregaron **en SPSS**. Cualquier cambio o ajuste de la base se hará a partir de la base maestra para mantener la consistencia entre bases. En el **Anexo 8** se presenta el requerimiento que se tomó como base para la entrega de las bases de datos.

