Longitudinal and Life Course Studies • vol xx • no xx • 1–15
 • Paper received November 2019/Revised April 2020/Published online April 2020
Online ISSN 1757-9597 • https://doi.org/10.1332/175795920X15858040679570

STUDY PROFILE

The Aguascalientes Longitudinal Study of Child Development: Baseline and First Results

Alfonso Miranda, alfonso.miranda@cide.edu
Economics Division, Centre for Economic Research and Teaching (CIDE) and Programme for Longitudinal Studies, Experiments and Surveys (PANEL), CIDE, Mexico, Institute for the Study of Labour (IZA) Germany

Osiel González Dávila, osiel.gonzalez@cide.edu
Programme for Longitudinal Studies, Experiments and Surveys (PANEL), CIDE, Mexico and National Council for Science and Technology (CONACYT) Mexico

Adriana Aguilar-Rodriguez, aaguilar@centrogeo.edu.mx
Centre for Research in Geospatial Information Science (CentroGeo), Programme for Longitudinal Studies, Experiments and Surveys (PANEL), CIDE and National Council for Science and Technology (CONACYT) Mexico

Antonio A. Aréchar, antonio.alonso@cide.edu
Programme for Longitudinal Studies, Experiments and Surveys (PANEL), CIDE and National Council for Science and Technology (CONACYT) Mexico

Daniel Zizumbo-Colunga, daniel.zizumbo@cide.edu
National Council for Science and Technology (CONACYT) Mexico and Department of Political Science, Vanderbilt University.

Yahaira Rodríguez-Martínez, yahaira.rodriguez@cide.edu
Programme for Longitudinal Studies, Experiments and Surveys (PANEL), CIDE, Mexico

Jaime Sainz-Santamaría, jaime.sainz@cide.edu
Public Administration Division, CIDE, Mexico and Programme for Longitudinal Studies, Experiments and Surveys (PANEL), CIDE, Mexico

This paper introduces the readership to the Aguascalientes Longitudinal Study of Child Development (EDNA) and presents the first descriptive results. EDNA is a prospective, multi-thematic, and multidisciplinary longitudinal study of the cohort of children that began first grade in August 2016 in the public schools of the state of Aguascalientes, Mexico. The sample contains a group of 1,000 pupils from 100 public schools who are representative of the study population. Recontact is planned to occur every two years for at least three waves. The baseline survey was conducted between 2017 and 2018. The study consists of an interview with the primary carer of the randomly chosen pupil, an interview with the pupil at school and an interview with the pupil's teacher at the premises of the Institute of Education of Aguascalientes. EDNA aims to identify and contribute to solving the problems faced by Mexican children to achieve healthy physical and intellectual development on their way to adult life.

Key words: longitudinal study • child development • Mexico • Aguascalientes

Key messages

- We introduce the baseline of the Aguascalientes Longitudinal Study of Child Development (EDNA).
- EDNA is a prospective, multi-thematic, multidisciplinary longitudinal study of child development.
- It follows a group of 1,000 children from 100 public primary schools who are a representative sample of the population.
- Main areas of study: (1) education and civic formation; (2) health and environment; (3) poverty and inequality.

To cite this article: Miranda, A., González Dávila, O., Aguilar-Rodriguez, A., Aréchar, A.A., Zizumbo-Colunga, D., Rodríguez-Martínez, Y. and Sainz-Santamaría, J. (2020) The Aguascalientes Longitudinal Study of Child Development: Baseline and First Results, *Longitudinal and Life Course Studies*, vol xx no xx, 1–15, DOI: 10.1332/175795920X15858040679570

EDNA's aims in brief

In developed countries, longitudinal studies of child development have been well established for several decades. Take, for example, the National Child Development Study (NCDS), the NICHD Study of Early Child Care, the Munich Longitudinal Study on the Genesis of Individual Competencies (LOGIC) and the Avon Longitudinal Study of Parents and Children (ALSPAC). In contrast, only a handful of projects exist in Latin America. Important pioneer studies include the aeioTU Longitudinal Study (Colombia) and the ELPI Longitudinal Survey of Early Life (Chile). In Mexico, only a few longitudinal studies have been conducted, with two worthy of mention: the Mexican Health and Ageing Study (MHAS), which followed 16,540 individuals aged 50 or over from 2001 to 2015 and is due to continue in 2021, and the Mexican Family Life Survey (MxFLS), which follows 35,000 individuals (Rubalcava and Teruel 2013, Wong et al, 2015). Neither focused on child development.

The Aguascalientes Longitudinal Study of Child Development (EDNA) is intended to bridge the gap. Aguascalientes is a small state in northern-central Mexico (see Figure 1). It has 11 counties or *municipios*, with a total population of 1.3 million (see Figure 2). The population is highly concentrated in the city of Aguascalientes, whose metropolitan area spans the capital *municipio* of Aguascalientes and the *municipio* of Jesús María. Nearly 877,000 live in the *municipio* of Aguascalientes, followed by 120,000 in Jesús María, 56,000 in Calvillo, and 53,000 in Rincón de Romos. The rest of the state has settlements with fewer than 50,000.

The study is supported by a strategic partnership between the Centre for Research and Teaching in Economics (CIDE) and the Aguascalientes Institute of Education (IEA), the local education authority. Funding was provided by CIDE and a grant by The Hewlett Foundation. The total budget of less than 65,000 US dollars, limited how ambitious the study could be in terms of sample size, but the funding allowed the research team to enter the partnership with absolute independence regarding the study's content and design. The partnership is important for: (1) establishing close collaboration with one of the most important stakeholders on the policy topics addressed by the study (the Aguascalientes government); (2) gaining access to the IEA's rich administrative records of schools and pupils, which are invaluable resources for designing and maintaining the study; and (3) obtaining assistance from the local authorities to initiate contact with schools and parents.

EDNA aims to identify, and contribute to solving, the main problems facing contemporary Mexican children in achieving healthy physical and intellectual development on their way to adult life. The methodological challenges are to generate data that allows causal inferences to be drawn about policy-relevant child development outcomes, and to contribute to designing, evaluating and improving public interventions in Mexico. The long-term project is to follow a sample of individuals from childhood to adulthood. The child is the focus of the study, so EDNA samples pupils in schools – more details on sampling design are given in

Figure 1: Aguascalientes state

Figure 2: *Municipios* (counties) of Aguascalientes: 001 Aguascalientes, 002 Asientos, 003 Calvillo, 004 Cosío, 005 Jesús María, 006 Pabellón de Arteaga, 007 Rincón de Romos, 008 San José de Gracia, 009 Tepezalá, 010 El Llano, 011 San Francisco de los Romo.

the section 'Sample design'. However, we understand that child development is a complex process that at minimum involves the child, their primary carer and their teacher. We therefore interview the sampled children's primary carers at home and their teachers at the premises of the IEA. Using these three elements together is one of the key features of EDNA.

This paper presents the baseline survey and first results, which will be a key contribution to future studies.

The social problems that affect Mexican children's development are multiple, complex and intertwined. For this reason, EDNA is a multi-thematic and multidisciplinary study with a life-course (or longitudinal) perspective. Life-course theory is a multidisciplinary methodological perspective that originated in around the 1920s – with important developments in the second half of the century (Elder et al, 2003). Life is seen as a succession of 'events', or a 'trajectory', that takes an individual from birth to death. According to this theoretical perspective, such a trajectory is key to understanding why an individual behaves the way they do at a given point in time and how the actions they take determine their future trajectory. By considering time as a dimension of analysis, the life-course perspective provides an opportunity to verify the premise that effects cannot precede causes, which is paramount to the scientific method. It is therefore possible to reject theories that are inconsistent with observed evidence. Observing an individual over time makes it possible to identify causal effects/mechanisms exploiting policy interventions or natural shocks that can be plausibly treated as exogenous to the study. For example, it would be possible to identify the causal effect of reducing teaching hours on children's educational

Table 1: Distribution of study population by *municipio*

	Pupils	Prop.
Aguascalientes	16,417	0.63
Asientos	1,111	0.04
Calvillo	1,103	0.04
Cosío	375	0.01
Jesús María	2,470	0.09
Pabellón de Arteaga	929	0.04
Rincón de Romos	1,189	0.05
San Francisco de los Romo	1,300	0.05
San José de Gracia	212	0.01
Tepezalá	545	0.02
El Llano	455	0.02
Total	26,106	1

outcomes using the 2019/20 coronavirus outbreak (COVID-19) as a natural shock that reduces the number of days schools remain open in the 2019/20 academic year. Indeed, through dynamics and interactions over time, individuals' socio-economic characteristics and the environment, longitudinal studies allow researchers to identify and separate age, period and cohort effects – or at least two of them. Such studies can also quantify contextual effects of changes over time in the social, physical, historical, and cultural environments on the life trajectory (Blau 1960, Davis et al, 1961, Lazarsfeld and Menzel, 1961, Hauser, 1970, Blalock, 1984, Robinson, 2009).

During their life, individuals go through different stages of development: from gestation to infancy and from youth to adulthood. In each stage, body and mind change, gaining and losing motor, cognitive, non-cognitive and reproductive abilities. Most life courses are dynamic and there are changes in direction and velocity at different points. Though each developmental stage lasts for more or less a given period, transitions from one stage to another vary from individual to individual. Furthermore, not all stages are hierarchically equal from a developmental point of view: there are critical points and breakpoints (for more on life-course theory see, for example, Elder, 1998).

Sample design

The study population is the set of all 26,106 children who started first grade in August 2016 at one of the 563 public primary schools in the state of Aguascalientes, Mexico.

Under a strict confidentiality agreement, the IEA provided us with a comprehensive list of primary schools in Aguascalientes and all children who had enrolled at each school by June 2016 to start first grade in August. This list of schools and enrolled children was used as a *frame* for selecting the sample. We hereafter call the number of pupils who were enrolled to enter first grade by June 2016 the 'expected size of the 2016 cohort'. The frame contains a measure of social marginality at the locality or settlement level, which is a four-level indicator of poverty generated by CONAPO (2012) (the National Population Council) based on the 2010 Census – hereafter the 'poverty index'.

We are interested in studying schools and pupils in poor communities and reflecting the diversity of the study population across the whole state of Aguascalientes, but we do not want to end up with a sample plagued with cell-size problems. By looking at the distribution of the study population by *municipio* and poverty condition (see

Table 2: Distribution of study population by marginality (poverty) level

	Pupils	Prop.
High	1,090	0.04
Medium	2,487	0.10
Low	7,862	0.30
Very low	14,667	0.56
Total	26,106	1

Table 3: Distribution of study population by selected group

Municipio	**Low poverty**		**High poverty**		**Total**	
	Pupils	**Prop.**	**Pupils**	**Prop.**	**Pupils**	**Prop.**
Aguascalientes	15,301	0.59	1,042	0.04	16,343	0.63
Other *municipios*	7,612	0.29	2,151	0.08	9,763	0.37
Total	22,913	0.88	3,193	0.12	26,106	1

Tables 1 and 2), we concluded that much of the study population lives in the *municipio* of Aguascalientes in localities with very low poverty as classified by CONAPO. The rest of the study population is quite geographically dispersed, with less than 15% living in localities with high or very high poverty levels. If we define strata using the 11×4 *municipio*-by-poverty table, we will end up with undesirably small cells. To avoid this, we decided to keep the *municipio* of Aguascalientes on its own and put all other *municipios* in a second group. For the poverty dimension, we decided to put the population living in localities with low and very low poverty levels in one group and localities with high and medium poverty levels in another group. This defines a 2×2 stratification table (see Table 3), with the disadvantage that there are only 1,042 pupils in the smallest cell – representing 4% of the study population. This is clearly unsatisfactory because, short of doing a census for that cell or a complex sample design, a random sample will produce a cell with only a few observations. To avoid this, we decided to combine the cells in the high poverty column of Table 3 to form a single group that represents 12% of the study population. With all these considerations in mind, we defined three strata:

> *Stratum 1:* Schools and pupils in localities with low or very low poverty in the *municipio* of Aguascalientes.
> *Stratum 2:* Schools and pupils in localities with low or very low poverty in any other *municipio* in the state.
> *Stratum 3:* Schools and pupils in localities with medium or high poverty anywhere in the state.

We expected this to work well for achieving our stratification objectives – they reflect the structure and diversity of the population and yet avoid problems with cell size. Beyond enrolment numbers, the frame did not have information that could be used to estimate the characteristics of the study population in each stratum. As we were working with a cohort that had not yet started primary education, there were no measures of previous attainment. Also, as far as we know, no previous surveys have been done on our study population. Therefore, we did not have any information that could be used as a measure of how the variance changed from stratum to stratum to implement an optimal sample design. As a consequence, we settled for using a

Table 4: Study population by stratum

	Pupils	Prop.
Stratum 1	15,141	0.58
Stratum 2	7,571	0.29
Stratum 3	3,394	0.13
Total	26,106	1

proportional allocation of the sample across strata. Proportional allocation is a popular choice, as it is more efficient than other sample allocation schemes when there is no prior information on the variance in each stratum (Lehtonen and Pahkinen 2004: 64–5). Table 4 shows how the study population is partitioned by our stratification scheme.

To determine the sample size, we started by calculating, under simple random sampling, the minimum number of observations that are required to perform inference for a difference of proportions at a significance level of $\alpha = 0.05$ and a maximum margin of error $E = 0.05$ using a binomial test and ignoring stratification,

$$n = \frac{Z^2_{1-\left(\frac{\alpha}{2}\right)} p\left(1-p\right)}{E^2}$$

where $Z^2_{1-\left(\frac{\alpha}{2}\right)}$ is the standard normal upper $100\left(1-\frac{\alpha}{2}\right)$ quantile, and $\lceil \cdot \rceil$ denotes the ceiling function. Setting $p = 0.5$ to consider the case of maximum variability, we obtain $n = 384$, which we used as a reference for the minimum acceptable sample size.

After considering logistics and funding limitations, we decided to aim for a representative sample of the study population of 1,000 pupils, which attains a maximum margin of error $E = 0.03$ for a difference in proportions at 5% of significance – ignoring stratification.

We decided to sample ten pupils per school for two reasons. First, attrition is to be expected in a longitudinal study and starting with fewer than ten pupils per school was thought to be risky as the sample could rapidly deteriorate in some schools. Second, having more than ten pupils per school would reduce the number of schools that could be included in the sample and potentially hinder our ability to do inference at the school level.

Due to the lack of previous knowledge, we anticipated a conservative unit non-response rate of 25%. As a consequence, we concluded that the number of attempted interviews that were needed to achieve our target sample was 1,334 pupils and 134 schools.

For selection, we followed a classic two-stage stratified cluster sampling design, with schools (PSUs) selected in the first stage and pupils (SSUs) selected in the second stage. Schools were drawn with a proportional-to-size (PPS) selection scheme – with replacement – implemented with systematic sampling to implicitly stratify for school shift (mornings/afternoons/extended hours). For the PPS, we used as a measure of size the expected size of the 2016 cohort, which we knew from the frame. Notice that for the PPS selection of schools to the sample, the measure of size only considers expected first-grade pupils, not pupils from other grades. Notice also that using a PPS selection scheme implies a varying probability of selection of schools in the first stage depending on their size as measured by our auxiliary indicator of size – that is,

Table 5: Final sample by stratum

	Pupils	Prop.
Stratum 1	779	0.57
Stratum 2	390	0.29
Stratum 3	195	0.14
Total	1,364	1

the expected size of the 2016 cohort. In particular, under PPS, larger schools were drawn into the sample with a higher probability than smaller schools.

After implementing PSU selection in the first stage, we ended up with 78 schools in stratum 1, 39 in stratum 2 and 18 in stratum 3 (accepting an increment in the number of schools from 134 to 135 due to rounding). However, to ensure that the smallest stratum had at least 20 schools, we decided to add two more schools into stratum 3. This meant that the number of attempted interviews needed to achieve our target sample increased to 1,370 and the number of schools to 137.

For the second stage, our frame contained a preliminary list of pupils enrolled in the schools selected in the first stage by June 2016. However, in initial approaches to authorities in the selected schools, we detected that many children who were enrolled in June 2016 had moved to another school during the first three months of the academic year. In most cases, such movements did not affect the total number of first-grade pupils in the school – when a child moved away, another was normally allowed to enter the school in their place. Therefore, the frame of pupils in EDNA's selected schools needed to be updated in the field before it could be used to draw our sample of pupils. This update became part of the initial contact protocol. Once the frame was up to date, ten pupils were drawn at random within each selected school with equal probability using systematic sampling to implicitly stratify classes. Schools in Aguascalientes vary in size. Small schools only have one first-grade class, while large schools have up to three. Class sizes range from 15 to 40 pupils, except in a small number of rural schools where there are fewer than ten first-grade pupils.

This is a classic two-stage PPS design in which the (unequal) selection probabilities of schools in the first stage are offset by the (unequal) selection probabilities of pupils in the second stage so that each pupil in the population has the same, non-negative, probability of entering the sample. That is, we have an equal probability sample (EPS) design. Table 5 shows the final issued sample: $n = 1,364$ pupils and 137 schools – we end up with an issued sample of $n = 1{,}364$ in 137 schools because there were some rural schools selected that had fewer than ten first-grade pupils belonging to the 2016 cohort.

Under the assumptions that we effectively reach our target, and that non-response is equally distributed across strata, the achieved sample size should be around $n = 1,023$ pupils in 100 schools, and we expect to obtain 584 completed interviews in stratum 1, 293 completed interviews in stratum 2, and 146 completed interviews in stratum 3. This means that, using a binomial test for a difference in proportions at 5% of significance, we expect a maximum margin of error E = 0.041 in stratum 1, E = 0.057 in stratum 2, and E = 0.081 in stratum 3. With these parameters, it is possible to attain a power of $1 - \beta = 0.99$ in stratum 1, $1 - \beta = 0.92$ in stratum 2, and $1 - \beta = 0.64$ in stratum 3 in a test for a difference in proportions of $d = 0.1$, where β represents the type II error in a binomial test for a difference in proportions at 5% of significance. However, if we were to test for a difference of $d = 0.12$ then we

would achieve a power of $1 - \beta = 0.8$, which is quite acceptable. Finally, at the school level, 100 observations are expected to attain a maximum margin of error E = 0.098.

Calculations a posteriori – using the final achieved sample – of the design effect relative to simple random sampling (DEFF) for the proportion of females in the study population give a DEFF = 0.99. Moreover, for the population mean of a continuous variable such as pupil's height, we obtain a DEFF = 1.03 for girls and DEFF = 1.00 for boys. From these results, we conclude that EDNA's sample design does a good job relative to SRS.

Once the sample of pupils and schools was determined, we initiated contact with primary carers.

Ethical issues and first contact protocol

EDNA complies with the strictest ethical standards. Every research topic included in the study has a clear social benefit and all questions pose minimal risk to research participants.

All interviewers received rigorous training, which included implementing safety protocols during fieldwork. All instruments were peer-reviewed and sensitive questions were carefully crafted to ensure minimal risk.

The IEA sent an advance letter to the principals of all EDNA-selected schools to explain the purpose of the study, instructing cooperation and introducing EDNA's research team.

In a first visit to the school, EDNA personnel talked to the principal and clarified doubts. At that point, the school's list of first-grade pupils – the school's frame – was checked and updated with the help of the principal and first-grade teachers. Once the school frame was updated, ten pupils were randomly selected with equal probability.

The list of selected pupils was given to the principal and first-grade teachers in a second visit to the school. At that point, EDNA's team asked for help to assemble primary carers for a meeting at the school's premises.

At the meetings, primary carers received information about the objectives, scope, methods, risks and benefits of the study. They were informed that participation was voluntary, that no financial compensation would be given and that all collected data would be treated with strict confidentiality and respect for privacy. At the end of the meeting, primary carers who decided to participate in the study gave contact details to EDNA's team, including their name, address and telephone number. Once this was done, we started the main fieldwork, interviewing primary carers of the chosen EDNA children in their homes.

For the primary carer interviews, participants were again informed of the risks and benefits of participating in the study and asked to sign an informed consent form. At this point, parental permission for the pupil's interview at school was requested in writing. In all cases, the assent of the participant was verbally requested before starting the interview. If at any point the interviewee expressed a desire to stop, the interview was immediately terminated. For analysis, all data is anonymised to guarantee strict confidentiality and respect for participants' privacy. To ensure maximum security in the handling of the information, all the principal investigators and research assistants signed a confidentiality agreement. Further, the original databases are not permitted to leave CIDE's premises. For further information on the ethical guidelines, see Miranda et al (2019).

Table 6: Nodal areas and domains in each questionnaire

Nodal area	Domains for mother's interview	Domains for child's interview	Domains for teacher's interview
Personal	Family	Socio-demographic	Socio-demographic
	Socio-demographic	Expectations	Professional profile
	Academic profile	Nutrition	Labour profile
	Labour profile	Family	
	Health and anthropometrics	Health and anthropometrics	Health and anthropometrics
Academic		Learning opportunities	Learning opportunities
			Use of technology
			Educational reform
Social environment		Bullying	Classroom climate
			Bullying
			Infrastructure
Self-reported	Cognitive skills (Raven)	Cognitive skills (Raven)	Cognitive skills (Raven)
	Growing mindset	Motivation	Grit
	Mental health	Personality (Field Berkeley Puppet Interview Big Five)	Mental health
	Big Five		Big Five
Civic participation	Civic commitment	Civic commitment	

Data collection methods and coverage

Our primary methods for data collection were face-to-face interviews and linkage to administrative records. While the primary carer interviews were conducted using pencil and paper, the pupil and teacher interviews were conducted using Computer-Assisted Personal Interviewing (CAPI).

Depending on the nature of the topic, subsections of the questionnaire – modules – contained a mixture of questions that were intended to be asked only once and questions that are intended to be asked in every wave. Dates of birth, sex and names are examples of variables that are only recorded once, whereas variables such as family composition and income are scheduled to appear, longitudinally, in every wave of the study.

Table 6 includes a summary of the key thematic areas included in each of the three interviews – pupil, primary carer and teacher.

The primary carer interview covered topics regarding family composition, education, anthropometrics, access to and use of health services, religion, citizenship and social capital, chronic diseases, water consumption, subjective well-being, work and income. We also collected a measure of cognitive ability (Raven), self-esteem (Rosenberg), mental health (PHQ9), parenting style, and personality (Big Five).

The pupil interview included questions regarding demographics, family and the socio-economic context in which the child is growing up, school performance, non-cognitive skills, physical and mental health, subjective well-being, citizenship and political attitudes, cognitive ability (Raven) and personality (Field Puppet Berkeley Interview).

Finally, the teacher interview included questions regarding demographics, professional and occupational profile, health, teaching practice, group structure, didactic planning, use of technology and the teacher's opinions about their pupils, school and classroom climate. There were also measures for cognitive ability (Raven), mental health (PHQ9) and personality (Big Five and Grit).

Interviewer profile, training and pilots

The team of interviewers consisted of four people with experience in survey collection, all holding a bachelor's degree – two trained as psychologists and two trained as social workers. Each interviewer attended compulsory training led by at least one senior member of EDNA's research team. Training covered use of electronic equipment and printed material, software use and questionnaire training, and confirmation that information was successfully recorded and loaded. In addition, interviewers were shown how to record notes, observations and/or comments about the interviews.

There were specific sessions for each instrument. An in-depth presentation of EDNA was followed by a detailed examination of the questionnaires. All the codes for the interview responses were reviewed. In addition, a survey-completion exercise was conducted. For the teacher interviews, which were self-administered, EDNA interviewers were taught how to train teachers on the spot to use the tablet to complete the questionnaire.

The primary caregiver questionnaire was piloted in several versions in October 2016 with help of a civil society organisation, *Bienestar e Integración Familiar* (BIFAM), which serves underprivileged children in the metropolitan area of Aguascalientes (*municipio* of Jesús María). Similarly, several versions of the child questionnaire were piloted between November 2017 and January 2018 at local schools that were not part of the EDNA sample. Finally, several versions of the teacher questionnaire were piloted between March and April 2018.

Fieldwork results

Thanks to the support of the IEA, there were no rejections at the school level. The main fieldwork for the primary carer interviews took place between January and August 2017. We completed 1,143 of the 1,364 planned interviews – a response rate of 84%. In the case of pupils, fieldwork started in February 2018 – when the children were in second grade and aged between seven and eight – and ended in June 2018. A total of 1,118 pupil interviews were completed out of the 1,143 originally planned – a response rate of 98%. Thanks to the IEA's administrative records, we were able to find all children who changed schools between the 2016/17 and 2018/19 school years and interview them. A household visit was attempted in cases where we detected that a child was absent from the school for more than two weeks. Finally, for teachers, fieldwork spanned May to September 2018, with 176 of 235 interviews completed – a response rate of 75%, which is a relatively high in comparison to similar studies that interview teachers. For instance, in the UK Millennium Cohort Study (MCS), the teachers' response rate was about 70%.

The final datasets contain 1,143 valid records of primary carers, 1,118 valid records of children, and 176 valid records of teachers. Merging primary carers and children results in a total of 1,074 valid primary carer–child pairs. Similarly, merging children

and teachers results in a total of 830 child–teacher valid pairs. Finally, merging all three datasets results in a total of 830 primary carer–child–teacher triads.

Frequency and panel maintenance

There will be an EDNA follow-up every two years for at least three waves. Fieldwork for the second wave is expected to start in September 2020, when the children will be in fifth grade and aged between nine and ten. Funding for the second wave is now secured.

The EDNA baseline survey asks primary carers to nominate three stable contact names and addresses. Contact details for the children's grandmothers were explicitly sought, as most of them are approaching retirement age and are unlikely to move.

Panel maintenance will also exploit the close relationship between CIDE and the IEA, which grants access to administrative records about pupils, teachers and schools. In particular, the IEA's administrative records include a unique national identification number that makes it possible to track each pupil in the population over time as long as the child remains in education. This means that the IEA knows which school each pupil in the population is attending at all times regardless of year-to-year movement between schools. This will be invaluable for tracking difficult cases and maintaining the quality of the study.

Policy and procedures for data access

All databases will be published on EDNA's website (www.cide-edna.org) on 20 December 2020. A teaching data set is currently publicly available. To have access, users are required to complete a registration form that is designed to keep track of data usage. Use of EDNA's data must be acknowledged by citation, indicating:

1. The title of the study, including the acronym (EDNA) and the year of implementation (2018 for the baseline survey, for instance).
2. The data set version.
3. The source and date of the download.

Descriptive statistics

Primary carer

A total of 1,128 respondents self-identified as a child's primary carer; the majority were born in Aguascalientes (79%), received no education beyond secondary school (74%) and are Catholic (92%). Average household size is 4.4 members (SD = 2.0).

Regarding economic activity, 90% had worked for a salary at some point in their lives but only 49% reported being in paid employment currently. Among those who were working, primary carers reported working on average 7.2 hours per day (SD = 3.1) and 4.9 days per week (SD = 1.6). The monthly salary was on average 4,336 Mexican pesos (SD = 4,728) – about US$217 – which represented 48% of total household income on average.

Regarding health, 61% of primary carers reported drinking alcohol in the last year while 67% declared that they had never smoked in their lives. Primary carers

weighed an average of 70.1 kg (SD = 15.9) and were 1.59 m tall (SD = 0.1). As a consequence, body mass index (BMI) was on average 27.8 (SD = 6.0). Nearly 67% were above BMI = 25, a conventional boundary for being overweight or obese. In terms of chronic diseases, 14% declared a diagnosis of hypertension and 6% a diagnosis of diabetes. Our findings indicate that most carers did no exercise, with less than 42% declaring that they exercised for at least 30 minutes three times per week in the last month. Nevertheless, 71% of the primary carers claimed to be in good health and 61% claimed to be very satisfied with life.

Pupils

We interviewed 1,118 pupils, of whom 51% were female. They were, on average, 7 years and 3 months old (SD = 0.49) at the time of the interview. Some 56% wanted to become a professional, 25% an employee and 7% an athlete.

On average, children were 127 cm tall (SD = 7.8) and weighed 27 kg (SD = 5.9). Using the WHO Child Growth Standards of BMI-for-age, we found that 2% of the children were underweight, 73% were normal weight, 14% were overweight and 11% were obese.

About one in five did not usually have breakfast before coming to school and almost half only had a light breakfast. Some 15% preferred soda (a carbonated sugary drink) over water, even though 97% said that water is healthier than soda.

About 57% of the children used electronic devices, mainly recreationally: 29% used mobiles, 35% used tablets and only 18% used laptops or computers. Extracurricular activities were varied: 80% reported watching TV after school, 42% doing sports, 65% reading and 84% doing homework. Only 8% worked outside their home for a salary.

The mother–child relationship is quite important. Around 75% said that their mother was the person helping with their homework – less than 19% were helped by their father.

About 27% declared having been smacked, 21% being punished with restricted TV and 24% being punished with restricted money. This makes physical violence the most frequent form of punishment ($p < 0.05$). Despite this, most children (76%) saw their family as closely united.

Children have strongly internalised gender stereotypes, with 67% believing that only men should work. Further, 76% believed that only women should clean the house, 86% that only women should cook and 68% that only women should take care of children. Boys and girls shared similar gender views.

Teachers

A total of 176 teachers were interviewed, of whom 90% were female. Most had a bachelor's degree (72%). In terms of experience, 22% said they had been in the profession for 6 to 10 years, 18% for 11 to 20 years and 30% for 21 to 40 years. The majority had a permanent position (55%), while 45% belonged to the Teacher Career Programme – a federal programme that grants salary supplements to teachers who perform well in an evaluation. Some 17% reported having more than five pupils with disciplinary problems in their class. Similarly, 16% had a class with one or more pupils with concentration problems, 36% had one or more visually impaired pupils and 37% taught at least one pupil with speech impediments.

Concerning school infrastructure, 36% thought natural lighting conditions in their classroom were inadequate and 18% thought their classroom was too small. Classroom ventilation is very important for teaching and learning, but 17% thought that their classroom was poorly ventilated.

Regarding learning opportunities, only 44% said they spend at least half of their time on teaching and learning activities. Other activities consume time: 65% said they spend more than 10% of their time maintaining discipline and 52% said they spend at least 10% of their time on administration.

In terms of teaching style, while 86% believed pupils should be allowed to think for themselves and 63% thought that pupils learn better if teachers support productive struggle in learning, only 57% agreed that their role is to facilitate pupils' research.

Conclusions

In this paper, we profiled the Aguascalientes Longitudinal Study of Child Development (EDNA). This is a prospective cohort study with recontact every two years. EDNA will be a very useful tool for policy makers, scholars and interested stakeholders.

Funding

The authors are grateful for funding received from the Cátedras CONACYT programme [project no. 874], The Hewlett Foundation [project no. 2013-8758], and CIDE [project FAI 12171077].

Acknowledgements

The usual disclaimers apply. We are grateful to the IEA for its support of EDNA. Particularly, we would like to thank Raúl Silva Pérezchica and María de Lourdes Carmona Aguiñaga; none of this work would have been possible without their enthusiasm and aid. We are also grateful to Hernán Bejarano, who contributed to the design of EDNA as a whole and the design of the mother and child questionnaires in particular. We thank Guillermina Jasso, from NYU, for her useful comments and encouragement, and Susie Jackson for copyediting the present paper.

Data availability statement

The EDNA research team takes responsibility for the integrity of the data and the accuracy of our analyses. All databases will be published on EDNA's website (www.cide-edna.org) on 20 December 2020. A teaching data set is currently public on EDNA's website.

Conflict of interest

The authors declare that there is no conflict of interest.

References

Blalock, H.M. (1984) Contextual-effects models: theoretical and methodological issues, *Annual Review of Sociology*, 10(1): 353–72. doi: 10.1146/annurev.so.10.080184.002033

Blau, P.M. (1960) Structural effects, *American Sociological Review*, 5(2): 178–93. doi: 10.2307/2092624

CONAPO (2012) *Índice de marginación por localidad 2010*, Benito Juárez, Mexico: Colección Índices Sociodemográficos, http://www.conapo.gob.mx/es/CONAPO/Indice_de_Marginacion_por_Localidad_2010

Davis, J.A., Spaeth, J.L. and Huson, C. (1961) A technique for analyzing the effects of group composition, *American Sociological Review*, 26(2): 215–25. doi: 10.2307/2089857

Elder, G.H., Jr (1998) The life course as developmental theory, *Child Development*, 69(1): 1–12. doi: 10.1111/j.1467-8624.1998.tb06128.x

Elder, G.H., Jr, Johnson, M.K. and Crosnoe, R. (2003) The emergence and development of life course theory, in J.T. Mortimer and M.J. Shanahan (eds) *Handbook of the Life Course, Handbooks of Sociology and Social Research*, Boston, MA: Springer, pp 3–20.

Hauser, R. (1970) Context and consex: a cautionary tale, *American Journal of Sociology*, 75(4, Part 2): 645–64. doi: 10.1086/224894

Lazarsfeld, P. and Menzel, H. (1961) On the relation between individual and collective properties, in A. Etzioni (ed) *Complex Organizations: A Sociological Reader*, New York: Holt, Rinehart & Winston, pp 422–40.

Lehtonen, R. and Pahkinen, E. (2004) *Practical Method for Design and Analyses of Complex Surveys* (2nd edn) Chichester: Wiley.

Miranda, A., González Dávila, O., Rodríguez Martínez, L. and Sainz Santamaría, J. (2019) Marco teórico general del estudio longitudinal del desarrollo de las niñas y los niños de Aguascalientes (EDNA), technical report, Mexico City: Centro de Investigación y Docencia Económicas (CIDE).

Robinson, W. (2009) Ecological correlations and the behavior of individuals, *International Journal of Epidemiology*, 38(2): 337–41. doi: 10.1093/ije/dyn357

Rubalcava, L. and Teruel, G. (2013) *Guía del usuario: Encuesta nacional sobre niveles de vida de los hogares tercera ronda*, Mexico City: Universidad Iberoamericana and CIDE at http://www.ennvih-mxfls.org/assets/guia_de_usuario_ennvih-3.pdf.

Wong, R., Michaels-Obregón, A., Palloni, A., Gutiérrez-Robledo, L., González-González, C., López-Ortega, M., Téllez-Rojo, M. and Mendoza-Alvarado, L. (2015) Progression of aging in Mexico: The Mexican Health and Aging Study (MHAS) 2012, *Salud Pública de México*, 57: S79–S89.