By how much has school participation declined as a result of the pandemic?

Martin Gustafsson
30 November 2021

This report was prompted by a problem in the analysis in the NIDS-CRAM Wave 5 report dealing with pandemic-related declines in school attendance. These declines were estimated to be in the region of 500,000. In that analysis, the pre-pandemic baseline estimate of dropping out, which drew from General Household Survey data, did not consider older learners, meaning the baseline estimate was too low. It was at first believed that correcting this would lead to a smaller impact on attendance. However, a re-analysis of the data pointed to a further problem. Once this was corrected, the loss in participation was found to be *greater* than the original 500,000. **My own estimate points to the number of non-returners between 2020 and 2021 being 910,000 higher than one would expect as a result of the pandemic. A colleague at Stellenbosch University arrived at a preliminary estimate of 600,000 after a separate re-analysis of the data. This is not as concerning as my 910,000, but also worse than the initial 500,000. The difference between my 910,000 and the lower 600,000 is not altogether surprising. The household data used for this work are not ideal for this purpose, and there are several ways of arriving at an estimate. What can be said with certainty, however, is that the published Wave 5 report figure of 500,000 is an *under*-estimate, not an *over*-estimate as was initially believed.** What can also be said with certainty is that **the loss has been the greatest at the primary level**. In fact, my analysis indicates that dropping out did not worsen in grades 10 to 12 during the pandemic. The original NIDS-CRAM report had indicated that the problem was worst at the secondary level.

The 910,000 loss figure is 8% of pre-pandemic attendance levels. Losses of this magnitude are also reflected in a couple of the few analyses currently available from the rest of the world. In Nigeria, it is estimated that attendance was 7% lower than normal after schools reopened, following pandemic-related closures. In the city of Charlotte in the United States, a similar loss has been reported, with that loss, like the South African one, being concentrated at the primary level.

I have worked closely with my Stellenbosch colleague, Dr Debra Shepherd, in revisiting the the estimate of the decline. While we may not arrive at exactly the same estimates, we have looked at each other's work carefully and are essentially in agreement that the original 500,000 was an under-estimate. Dr Shepherd is in the process of producing a separate report, which will discuss the data issues in more depth than the current one.

It must be emphasised that the NIDS-CRAM household data are far from ideal for this kind of analysis. To illustrate, the loss of 910,000 learners is extrapolated from telephonic responses of just 8,157 more or less randomly selected adults across the country, of whom just 476 said learners had not returned to school. Apart from the usual issue of sample-related confidence intervals, there is the risk that respondents did not understand the interviewer properly, especially given a confusing context where schools were implementing rotational attendance rules and that even attending learners were not experiencing normal schooling.

The pandemic has brought to the fore the importance of rapidly fixing problems with the manner in which the SA-SAMS system, used by around 90% of schools, feeds into provincial and national databases. Specifically, detailed learner attendance statistics should be more readily available. The SA-SAMS system is designed to capture attendance, and the data have been found to be of a high quality in at least one province, Limpopo. Data quality across the other eight provinces remains unclear. Myself and others are currently working on sourcing and analysing attendance data derived from SA-SAMS.

What the original NIDS-CRAM Wave 5 report of July 2021 found

The original report, by Shepherd and Moholwane (2021) and published in July 2021, concluded that there was “an increase in the number of absent learners of approximately 400,000-500,000 when compared to ‘normal’ times” (from the executive summary).

The report moreover explained that its point of departure that before the pandemic, specifically in 2018, around 231,233 learners would not return to school between one year and the next (p. 11). The source for this was the 2018 General Household Survey. However, this figure reflects the situation for learners aged 7 to 17 only. What is not taken into account is that by April of any recent pre-pandemic school year, 9% of learners are aged 18 and above. This is from analysis of the DBE’s national learner database. The percentage applies to April, because the NIDS-CRAM Wave 5 data were collected in April, and the percentage would change slightly according to the month of the year, as the birthdays of learners occur. One can expect high levels of dropping out among these older 9% of learners. However, the problem is that the GHS, unlike the NIDS-CRAM data, does not ask whether learners left school in the last year. It simply asks whether household members are currently attending school or not.

A reasonable estimate of learners dropping out per year before the pandemic

As will become clearer below, it is advantageous to use a household dataset from before the pandemic as a point of comparison. This is because the NIDS-CRAM data is household data. Using DBE data collected from schools would be an alternative, but then the problem arises that there is no information on certain household characteristics, making comparison to NIDS-CRAM statistics difficult.

The question is whether there is a way to use the GHS data to estimate how many learners leave school between one year and the next, even though the GHS does not ask this question. A key contribution of the current report is a proposal for doing this.

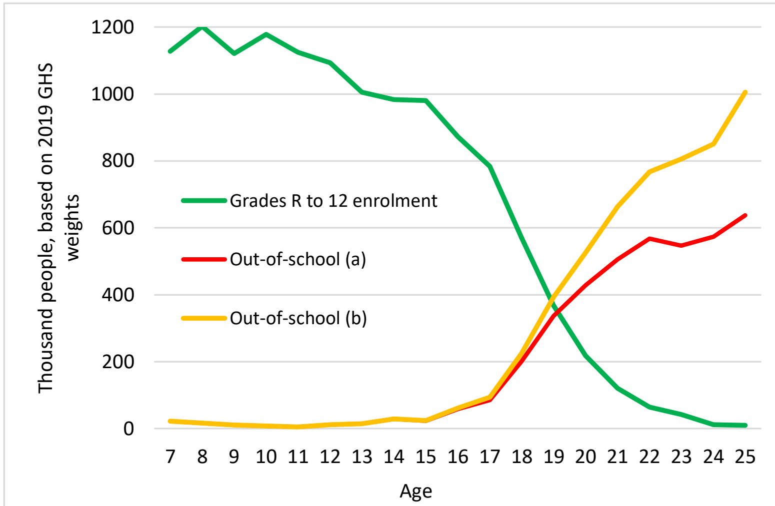
The 2019 GHS data are used, and not the 2018 data, as this more recent pre-pandemic dataset recently became available.

Using the 2019 GHS data, I conclude that around 370,000 learners left the schooling system per year before the pandemic. This considers only departures from grades 1 to 11. It is assumed that someone leaving Grade 12 is not a dropout, even if that person did not pass the National Senior Certificate examinations.

How one arrives at 370,000 pre-pandemic non-returners using the 2019 GHS

Figure 1 below shows the number of school-based learners per age in the 2019 GHS. It also provides two sets of estimates of the out-of-school, by age. Curve (a) assumes anyone who is not enrolled in *any* education institution *and* who is not earning a wage, is 'out-of-school'. Curve (b) provides higher figures as wage-earners are counted as the 'out-of-school'.

Figure 1: Out-of-school and enrolments by age

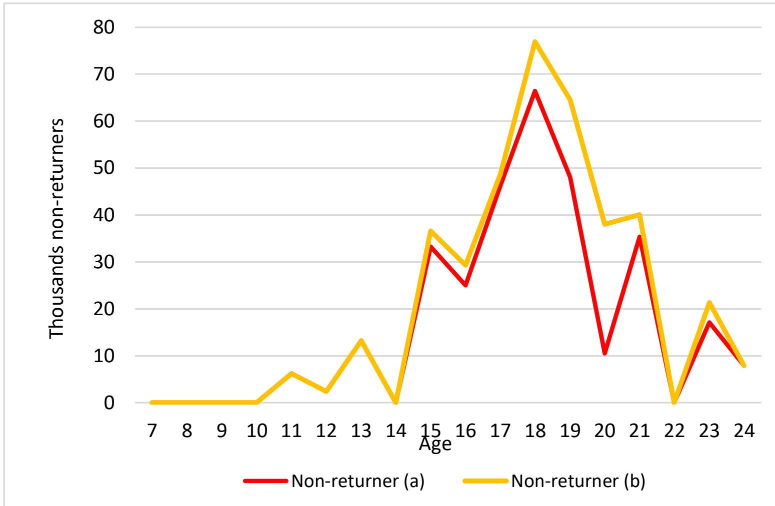


Apart from (a) and (b), a third calculation was also undertaken, which was like (b) but with the assumption that someone in adult education should also be considered out-of-school. This third approach yielded results which were almost indistinguishable from (b), and is thus not used in the discussion that follows.

An age-specific dropout rate is calculated, using just the statistics behind Figure 1. How this is done is explained with reference to those aged 18 in 2019. The number of people not in education aged 18 in 2019 is 224,916, according to the GHS weights – curve (b) is used here. For age 19 it is around 392,624. There is thus an increase per year for those aged 18 in the base year of 167,708, assuming that patterns do not change much over time. The number of 18-year-olds enrolled in grades R to 12 in a school (not the equivalent level in, say, a college) in 2019 was 570,338 – the green curve. The percentage of 18-year-old school learners becoming out-of-school the next year, assuming the next year is a normal non-pandemic year, is thus assumed to be 29% (167,708 over 570,338).

The next step was to randomly make a percentage of 18-year-olds enrolled in grades 1 to 11 in the 2019 GHS drop out. Exactly how this was done is shown in the code appearing in the appendix. Instead of the 29% referred to above, 23% was used. This 23% is what one obtains if one does the calculation using only households with members aged 7 to 17. The analysis needed to be limited to such households, for the appropriate comparison with the NIDS-CRAM data to be made. This compromises the explanation here a bit, but is the route to take if ultimately the aim is to compare to NIDS-CRAM. If one multiplies 23% by the number of 18-year-olds *in just grades R to 11,* one arrives at around 78,000 non-returners. This 78,000 for 18-year-olds is reflected in curve (b) in Figure 2 below. If such figures across ages 7 to 24 are added, the total one obtains is around 390,000. Each run of the method produces slightly different statistics as the GHS contains not individuals in the population, but sampled individuals who each represent several hundred people in the population. In the method described here, sampled individuals were not divisible.

Figure 2: GHS 2019 non-returners by age



The problem with the 390,000 figure, however, is that GHS weights exaggerate the size of the population slightly. Comparison of GHS weight totals for GHS learners to official enrolment figures indicates that actual figures are around 95% of GHS totals. The 390,000 must thus be reduced to 370,000.

A figure of 370,000 roughly agrees with what we know about non-completion of schooling in the country. In recent years, around 400,000 learners a year leave school before entering Grade 12. Around 30,000 of those who never enter Grade 12 move into a TVET college each year, though the exact number, and the associated delays, are not well-documented. A further 30,000 enter ABET Level 4 training, which is the equivalent of Grade 12[1]. That leaves around 340,000 youths, not very far from the 370,000.

The method behind the 910,000 attendance decline figure

The next step was compare GHS patterns to NIDS-CRAM Wave 5 patterns, and to arrive at a conclusion around the reduction in enrolments associated with the pandemic.

The discussion begins with the NIDS-CRAM data.

The critical question in the Wave 5 survey is the following: 'For learners in your household who have not yet returned to school, what grades are they doing?' The respondent can provide up to seven 'Yes' responses, each attached to a grade in the range R to 12. This question is only asked to adults who are in households where there are household members aged 7 to 17. An obvious risk with the question lies in the fact that it is not completely clear if grade refers to the grade of the learner in the previous year, or in the current year.

The point of reference for having returned is fairly clearly the 2020 school year, as shortly before the question mentioned above, another question is asked: 'How many young people in your household were attending school (Grade R-12) before schools closed in March last year?'.

[1] Gustafsson, 2020.

The respondent is thus likely to compare attendance in the current year with attendance in the previous year.

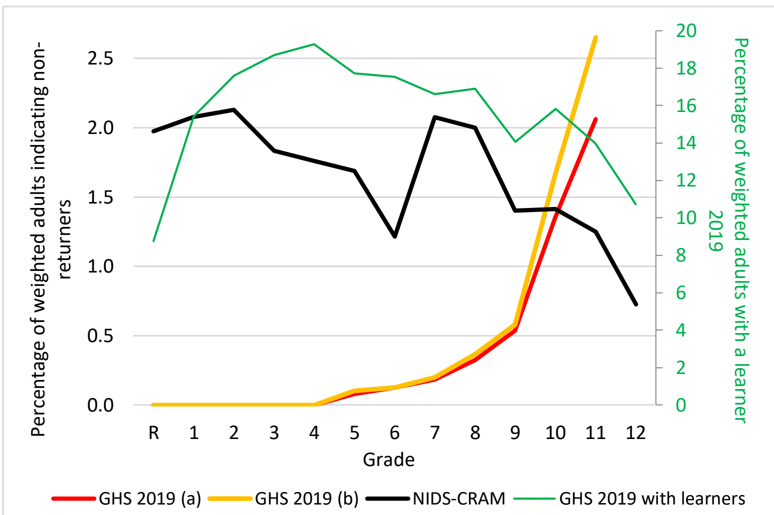
The left-hand vertical axis of Figure 3 below refers to the percentage of adults aged 18 and above in the population, *not to household heads* (though respondents may happen to be household heads). The NIDS-CRAM sample is essentially a random sample of South African adults. The metadata advises the following (Ingle *et al*, 2021: 9):

> …when discussing employment losses, a researcher should note: "These results are for a broadly representative sample of South African adults from 2017, who were re-interviewed in 2020 for NIDS-CRAM."

This clearly makes it possible for two or more NIDS-CRAM respondents to be from the same household.

Furthermore, the denominator for the left-hand vertical axis is adults from households where there are members aged 7 to 17. A household with just a 19-year-old Grade 11 learner would thus not be included in the analysis, for instance.

Figure 3: Comparing GHS 2019 to NIDS-CRAM 2021



What is immediately striking about the NIDS-CRAM curve in Figure 3 is that 1.5% to 2.0% of adults say someone in the household at the primary level has not returned to school. Under normal circumstances, this percentage would be extremely close to zero. It would have to be close to zero, or historical attendance rates which are around 98%, *and do not decline as age increases*, would not be possible – see for instance Department of Basic Education (2019: 6).

The red and orange GHS curves in Figure 3 draw from the GHS analysis described above, and mimic as closely as possible the methodology behind the NIDS-CRAM curve. To achieve this, first all households which did not have household members in the age range 7 to 17 were discarded. Secondly, all adults aged 18 and above from such households in the GHS were selected as units of the analysis. Clearly, a household could have more than one selected adult. Thirdly, if for instance it had been found, through the random process described previously, that a Grade 10 learner in 2019 had become a non-returner the next year, and if in the same household there was another Grade 10 learner who was a *returner*, then the assumed response for the other learner was 'non-returner', not 'returner'. This mimics the NIDS-CRAM data, where a respondent would say 'Yes' to Grade 10 if *any* Grade 10 learner in the household had not returned, regardless of whether there was *another* Grade 10 learner who *had* returned.

How common is it find two learners in the same household and grade in the GHS? Is is fairly common, and probably more common than one may think. For any single grade, in the case of around 6% households there is more than one learner in that grade. This 6% statistic does not vary much by grade.

It was thus assumed that in both the NIDS-CRAM data and the GHS data, if one learner in a grade had not returned, the same applied to the other learner. The latter may of course not be true. If it was true half of the time, then the estimate of the number non-returners would be around 3% higher than it should be.

The last thing that must be explained is the green curve in Figure 3, which should be read against the *right*-hand vertical axis. This curve indicates, for instance, that around 16% of South African adults living in households with members aged 7 to 17 would respond 'Yes' to the question of whether there was a Grade 6 learner in the household in 2019. The green curve draws from the GHS data.

If we know the actual enrolment in a grade, then the percentages represented in Figure 3 can be used to deduce the decline in attendance in that grade as a result of the pandemic. Grade 7 is used to illustrate the calculation. There were around 1,018,000 Grade 7 learners in 2019, according to the DBE's *2019 School Realities* publication (this includes public and independent schools). The green curve in Figure 3 reflects 16.6% for Grade 7. This 16.6% can be said to represent the 1,018,000 Grade 7 learners, because we make the reasonable assumption that all learners are in households that have adults, and the green curve represents responses from all adults (at least within households with members aged 7 to 17). The decline in the Grade 7 attendance level is represented by the vertical gap between the black and orange curves. For Grade 7 this would be 2.1% minus 0.2%, or 1.9. If we divide 1.9 by 16.6, we obtain 11.4%. The decline in Grade 7 attendance resulting from the pandemic was 11.4% of 1,018,000, or 115,000.

The orange curve (b) is used, meaning it is assumed that a NIDS-CRAM respondent considers someone who left, say, Grade 10 and found work to be a 'non-returner'. It is clear from Figure 3 that whether one uses the (a) or (b) curve makes hardly any difference.

One key assumption made is that the Grade 7 percentages would not be that different if households with *any* learners, and not just households with members aged 7 to 17, were considered.

Repeating the Grade 7 calculation across grades 1 to 11 produces a total of 938,000. The total becomes 1,121,000 if Grade R is included. The grades 10 and 11 results are negative, so these were changed to zero. Non-returning learners is in fact worse *before* the pandemic than during the pandemic for these grades – the black curve lies below the orange curve.

The 938,000 figure is an over-estimate insofar as it assumes that where one learner in the household had not returned, any other learners from the same household and in the same grade would also not have returned. If we assume that this is true only half of the time, and that when there is not just one learner per household and grade, the number is always two, then we should adjust 938,000 by 100 over 103. This gives a total of around 910,000. This 910,000 is 8% of the pre-pandemic grades 1 to 11 enrolment level.

In these calculations, the Grade 12 point for NIDS-CRAM has been ignored. Even if one were to take it into account in some way, this would not change the finding that the enrolment declines are concentrated *at the primary level.* Of the 910,000, the primary grades 1 to 7 level accounts for 758,000 of the lost learners.

The original Wave 5 NIDS-CRAM report had indicated that the drop in attendance was worst at the secondary level. The re-analysis of the data described here makes it clear that the problem was worst at the primary level.

It should be emphasised that the total number of NIDS-CRAM respondents behind the denominator of the black curve in Figure 3 is just 8,157. The numerator, in the sense of respondents saying they had non-returners in some grade, is just 476 people. In addition, adults were interviewed telephonically, which has limitations relative to face-to-face interviews, and the sample design is moreover particularly complex. The data are clearly only useful for drawing very broad conclusions, such as that the decline in enrolments is around 900,000, and could be in the range of 700,000 to 1.1 million. Such a wide range is driven not just by the sample size, but by the possibility that telephonic responses are not strictly in line with the question. In particular, it is very possible that some respondents answered 'Yes' to 'have not yet returned to school' if learners in the household were attending irregularly, in the context of the widespread use of rotational attendance rules applied by schools to reduce over-crowding. Irregular attendance is of course not the same as having dropped out of school entirely.

What about enrolment trends and the situation in other countries?

A recently released report compared *enrolments*, which is not the same as *attendance*, between the start of the 2020 and 2021 school years (Department of Basic Education, 2021). That report found enrolments declining by just 46,000. However, like the NIDS-CRAM data, that report found the largest loss to be at the primary level. In fact, at the Grade 12 level, enrolment is particularly high in 2021, and 20% higher than it was at the start of 2020.

It is quite possible that the pandemic caused a relatively small decline in enrolments, but a very large decline in actual attendance. Many learners may have been enrolled at the start of 2021, but then essentially not participated in schooling during the year.

The 8% decline in enrolment referred to above is not that unusual in the global context. There is still little analysis of the phenomenon in other countries, particularly developing countries. But a study from Nigeria, using data very similar to the NIDS-CRAM data, points to a decline in school attendance, after schools reopened, of 7% (Dessy *et al*, 2021: 24). In the city of Charlotte, in North Carolina in the United States, primary-level chronic absenteeism is reported to have risen by 7 percentage points, from 11% to 18% of enrolments, as result of the pandemic (Jordan, 2021: 2). In Charlotte, this deterioration has been far less serious at the secondary level. This is similar to what the NIDS-CRAM data reveal.

Efforts to improve the monitoring of learner attendance

The ongoing monitoring of learner attendance is clearly very important in the context of the pandemic. This monitoring would be vastly improved if the feed from the school-level SA-SAMS system into provincial and national data warehouses were improved. This is not that difficult to achieve, and is currently being explored in the DBE. The SA-SAMS system is used by over 97% of schools in six provinces, the figures being a lower 94% in KwaZulu-Natal, 86% in Gauteng, and 0% in Western Cape (the latter uses a different system)[2]. What is not known yet is the quality of the daily learner attendance data in SA-SAMS in most provinces. However,

[2] Department of Basic Education, 2020: 57.

this has been found to be extremely good in Limpopo, where attendance data were used to establish that menstruation cycles impact on female attendance (Van Biljon and Burger, 2019).

****

References

Department of Basic Education (2019). *General Household Survey (GHS): Focus on schooling 2018*. Pretoria.

Department of Basic Education (2021). *Impacts of the COVID-19 pandemic on school enrolments*. Pretoria.

Dessy, S., Gninafon, H., Tiberti, L. & Tiberti, M. (2021). *COVID-19 and children's school resilience: Evidence from Nigeria*. Washington: World Bank.

Gustafsson, M. (2020). *Broadening second-chance Matric opportunities in the context of the COVID-19 pandemic*. Cape Town: Youth Capital.

Ingle, K., Brophy, T. & Daniels, R. (2021). *Panel user manual: Release July 2021.* Cape Town: University of Cape Town.

Jordan, P.W. (2021). *Present danger: Solving the deepening student absenteeism crisis.* Washington: Georgetown University.

Shepherd, D. & Mohohlwane, M. (2021). *The impact of COVID-19 in education - more than a year of disruption*. Stellenbosch: NIDS-CRAM.

Van Biljon, C. & Burger, C. (2019). *The period effect: the effect of menstruation on absenteeism of school girls in Limpopo*. Stellenbosch: University of Stellenbosch.

Annexure: Stata code used

```
* ARRIVING AT 370,000 PRE-PANDEMIC DROPOUTS PER YEAR

use "C:\My Documents\Resources (Data)\Statistics South Africa (StatsSA)\General Household Survey\2019\GHS-2019-PERSON_F1.dta", clear
keep uqnr age EDU_ATTEND EDU_EDUI EDU_GRDE LAB_WGE person_wgt
save "C:\My Documents\Numbercrunching\GHS\temp0.dta", replace // temp0.dta is all the data needed
use "C:\My Documents\Numbercrunching\GHS\temp0.dta", clear
gen temp = 1 if age>=7 & age<=17
by uqnr, sort: egen haschild = max(temp)
keep if haschild==1
contract uqnr
drop _freq
save "C:\My Documents\Numbercrunching\GHS\temp1.dta", replace // temp1.dta is the identifiers of those households containing members aged 7 to 17
use "C:\My Documents\Numbercrunching\GHS\temp0.dta", clear
gen learners = person_wgt if EDU_GRDE>=0 & EDU_GRDE<=12 & EDU_EDUI==2
tabstat learners, stat(sum) format(%8.0f) // 13.7m
tabstat learners if age>=7 & age<=25, stat(sum) format(%8.0f) by(age) // Used for Figure 1 graph
tabstat learners if age>=7 & age<=25 & EDU_GRDE>=0 & EDU_GRDE<=11, stat(sum) format(%8.0f) by(age) // Used for Figure 1 graph
gen nonreturn1 = person_wgt if EDU_ATTEND==2 & LAB_WGE!=1
gen nonreturn2 = person_wgt if EDU_ATTEND==2
gen nonreturn3 = person_wgt if EDU_ATTEND==2 | EDU_EDUI==3
tabstat nonreturn* if age>=7 & age<=25, by(age) stat(sum) // Used for Figure 1 graph
merge m:1 uqnr using "C:\My Documents\Numbercrunching\GHS\temp1.dta"
keep if _merge==3
drop _merge
collapse (sum) nonreturn* learners if age>=7 & age<=25, by(age)
save "C:\My Documents\Numbercrunching\GHS\temp2.dta", replace
use "C:\My Documents\Numbercrunching\GHS\temp2.dta", clear
keep age nonreturn*
forvalues i = 1 / 3 {
  rename nonreturn`i' nonreturn`i'next
}
replace age = age - 1
merge 1:1 age using "C:\My Documents\Numbercrunching\GHS\temp2.dta"
drop _merge
forvalues i = 1 / 3 {
  gen dropout`i' = (nonreturn`i'next - nonreturn`i') / learners
}
```

```
keep if age>=7 & age<=24
save "C:\My Documents\Numbercrunching\GHS\temp2.dta", replace // temp2.dta is a small table with age-specific 'dropout rates'
use "C:\My Documents\Numbercrunching\GHS\temp0.dta", clear
merge m:1 age using "C:\My Documents\Numbercrunching\GHS\temp2.dta"
drop _merge
keep if EDU_GRDE>=0 & EDU_GRDE<=11 & EDU_EDUI==2
save "C:\My Documents\Numbercrunching\GHS\temp3.dta", replace // temp3.dta has an obs for every grades R to 11 learner from temp1.dta
forvalues a = 7 / 24 {
 use "C:\My Documents\Numbercrunching\GHS\temp3.dta", clear
 keep if age==`a'
 gen myrand = uniform()
 sort myrand
 gen running = sum(person_wgt)
 egen tot = sum(person_wgt)
 gen prop = running / tot
 forvalues i = 1 / 3 {
  gen innext`i' = cond(prop<dropout`i', 0, 1)
 }
 if `a'!=7 {
  append using "C:\My Documents\Numbercrunching\GHS\temp4.dta"
 }
 save "C:\My Documents\Numbercrunching\GHS\temp4.dta", replace // temp4.dta is temp3.dta with simulated dropping out added, and just ages 7 to 24 retained
}
use "C:\My Documents\Numbercrunching\GHS\temp4.dta", clear
forvalues i = 1 / 3 {
 gen outnext`i' = cond(innext`i'==1, 0, person_wgt)
}
merge m:1 uqnr using "C:\My Documents\Numbercrunching\GHS\temp1.dta"
tabstat outnext* learners, stat(sum) by(age) // Used for Figure 2 graph
tabstat outnext2 if age==18, stat(sum) // Around 77,000
tabstat outnext2, stat(sum) // Repeated run provides figure of around 390,000. 95% of this is taken, to deal with GHS over-weighting, producing the 370,000
tabstat outnext2 if _merge==3, stat(sum) // Repeated run provides figure of around 315,000

* GETTING THE NECESSARY NIDS-CRAM STATISTICS

use "C:\My Documents\Resources (Data)\SALDRU\CRAM Wave 5\NIDS-CRAM_Wave5_Anon_V1.0.0.dta", clear
merge 1:1 pid using "C:\My Documents\Resources (Data)\SALDRU\CRAM Wave 5\derived_NIDS-CRAM_Wave5_Anon_V1.0.0.dta"
drop _merge
merge 1:1 pid using "C:\My Documents\Resources (Data)\SALDRU\CRAM Wave 5\Link_File_NIDS-CRAM_Wave5_Anon_V1.0.0.dta"
drop _merge
```

```
recode w5_nc_no7to17res (0=0)(1/11=1)(.=-1), gen(childstatus)
forvalues g = 0 / 13 {
  gen noattp`g' = 0
  forvalues w = 1 / 7 {
    recode w5_nc_edunoatt2021lev`w' (-9/-8=13), gen(temp)
    replace noattp`g' = noattp`g' + 1 if temp==`g'
      drop temp
  }
}
tabstat noattp* if childstatus==1 [aweight = w5_nc_pweight_s], col(stat)  // Used for Figure 3 graph
count // 8,157
egen noattptot = rowtotal(noattp*)
count if noattptot!=0 // 476

* PRODUCING COMPATIBLE STATISTICS USING GHS 2019

use "C:\My Documents\Numbercrunching\GHS\temp4.dta", clear
merge m:1 uqnr using "C:\My Documents\Numbercrunching\GHS\temp1.dta"
keep if _merge==3
drop _merge
by uqnr EDU_GRDE, sort: egen tempcountl = count(_n)
egen temphhgrtag = tag(uqnr EDU_GRDE)
codebook tempcountl if temphhgrtag==1 // If not 1, then nearly all 2
gen tempnot1 = cond(tempcountl!=1, 1, 0)
tabstat tempnot1 if temphhgrtag==1, by(EDU_GRDE) // Around 6%
drop temp*
forvalues g = 0 / 11 {
  forvalues i = 1 / 3 {
    gen gr`g'_`i' = cond(EDU_GRDE==`g' & innext`i'==0, 1, 0)
  }
}
collapse (max) gr*, by(uqnr) // The use of max means that where in the same household a Grade 10 learner is a non-returner while another is a returner, both are counted as
      non-returners, to mimic the NIDS-CRAM data.
save "C:\My Documents\Numbercrunching\GHS\temp5.dta", replace // temp5.dta one obs per household, and grade-specific variables on whether a non-returner
use "C:\My Documents\Numbercrunching\GHS\temp0.dta", clear
merge m:1 uqnr using "C:\My Documents\Numbercrunching\GHS\temp1.dta"
keep if _merge==3
drop _merge
keep if age>=18 & age<=120
merge m:1 uqnr using "C:\My Documents\Numbercrunching\GHS\temp5.dta"
```

```
drop if _merge==2
drop _merge
forvalues g = 1 / 11 {
  forvalues i = 1 / 3 {
    replace gr`g'_`i' = 0 if gr`g'_`i'==.
  }
}
tabstat gr*_1 [aweight = person_wgt], col(stat) // Used for Figure 3 graph
tabstat gr*_2 [aweight = person_wgt], col(stat) // Used for Figure 3 graph
use "C:\My Documents\Numbercrunching\GHS\temp0.dta", clear
forvalues g = 0 / 12 {
  gen gr`g' = cond(EDU_GRDE==`g' & EDU_EDUI==2, 1, 0)
}
collapse (max) gr*, by(uqnr)
save "C:\My Documents\Numbercrunching\GHS\temp6.dta", replace // temp6.dta one obs per household, and grade-specific variables on whether a learner
use "C:\My Documents\Numbercrunching\GHS\temp0.dta", clear
merge m:1 uqnr using "C:\My Documents\Numbercrunching\GHS\temp1.dta"
keep if _merge==3
drop _merge
keep if age>=18 & age<=120
merge m:1 uqnr using "C:\My Documents\Numbercrunching\GHS\temp6.dta"
drop if _merge==2
drop _merge
forvalues g = 1 / 12 {
  replace gr`g' = 0 if gr`g'==.
}
tabstat gr* [aweight = person_wgt], col(stat) // Used for Figure 3 graph
```