**END-OF-TERM SYSTEM**

**(EOTS)**

**MHEC Internal Validation Code**

**STATA Syntax Code**

**Posted 10/03/2019**

# **Data Validation Syntax**

The following Stata code is used to validate all file submissions. Additional validation testing may be completed contingent upon the results of this baseline code. New validation steps may be added over time as the scope of the collection changes.

\*file population\*

\*compare totals for each tab to prior years\*

\*compare sic to fice for correct sic construction\*

table fice

table sic

\*count unique populations\*

\*first instance of each unit record within an fice is marked as unique, additional instances as duplciate\*

\*add to excel tab and cross check against prior year totals\*

\*eots should be one row per idn per fice, duplicates are errors\*

\*counts should be the same on idn, tssn, tssn2\*

sort fice idn

egen idn\_unique=tag(fice idn)

sort fice tssn

egen tssn\_unique=tag(fice tssn)

sort fice tssn2

egen tssn2\_unique=tag(fice tssn2)

label define counts 1 "unique" 0 "duplicate"

label values idn\_unique tssn\_unique tssn2\_unique counts

tab idn\_unique

tab tssn\_unique

tab tssn2\_unique

\*validate FICE-SIC alignment\*

gen sicfice=.

replace sicfice=1 if fice==143600& sic==260700

replace sicfice=1 if fice==205700& sic==110100

replace sicfice=1 if fice==205800& sic==110200

replace sicfice=1 if fice==206100& sic==111300

replace sicfice=1 if fice==206200& sic==120600

replace sicfice=1 if fice==206300& sic==111250

replace sicfice=1 if fice==206400& sic==111000

replace sicfice=1 if fice==206500& sic==263300

replace sicfice=1 if fice==206700& sic==261200

replace sicfice=1 if fice==206800& sic==121400

replace sicfice=1 if fice==207100& sic==111700

replace sicfice=1 if fice==207200& sic==121800

replace sicfice=1 if fice==207300& sic==262000

replace sicfice=1 if fice==207400& sic==112100

replace sicfice=1 if fice==207500& sic==112200

replace sicfice=1 if fice==207600& sic==262300

replace sicfice=1 if fice==207700& sic==262500

replace sicfice=1 if fice==207800& sic==262600

replace sicfice=1 if fice==208000& sic==262700

replace sicfice=1 if fice==208300& sic==143000

replace sicfice=1 if fice==208600& sic==263100

replace sicfice=1 if fice==208900& sic==113600

replace sicfice=1 if fice==209100& sic==123900

replace sicfice=1 if fice==209200& sic==263700

replace sicfice=1 if fice==209500& sic==154000

replace sicfice=1 if fice==209900& sic==124200

replace sicfice=1 if fice==210200& sic==124400

replace sicfice=1 if fice==210300& sic==124700

replace sicfice=1 if fice==210400& sic==124500

replace sicfice=1 if fice==210500& sic==124600

replace sicfice=1 if fice==210600& sic==124800

replace sicfice=1 if fice==210700& sic==265000

replace sicfice=1 if fice==210800& sic==265200

replace sicfice=1 if fice==210900& sic==265400

replace sicfice=1 if fice==465000& sic==111100

replace sicfice=1 if fice==691100& sic==112970

replace sicfice=1 if fice==817500& sic==112400

replace sicfice=1 if fice==830800& sic==110900

replace sicfice=1 if fice==1001400& sic==111900

replace sicfice=1 if fice==1164400& sic==124900

replace sicfice=1 if fice==2073900& sic==115470

replace sicfice=1 if fice==3100700& sic==110770

tab sicfice

\*Step 2: Duplicate Check\*

\*this must be done first as variables will be constructed throughout validation that will

\*alter results when looking for duplicates across all variables\*

sort fice idn

\*dups on all variables\*

duplicates tag, generate(dups) /\*duplicates in data set, data set duplicates on all elements are errors\*/

tab dups

\*dups on ssn only\*

\*validate counts are same on idn, tssn and tssn2 to make sure idn is correctly converted\*

\*id may appear multiple times if student concurrent enrolled in more than one FICE\*

duplicates tag idn, generate (dupssn) /\*not an error, student attending multiple fice\*/

duplicates tag tssn, generate (duptssn)/\*not an error, student attending multiple fice\*/

duplicates tag tssn2, generate (duptssn2)/\*not an error, student attending multiple fice\*/

tab dupssn

tab duptssn

tab duptssn2

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*validate individual data elements\*

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*valid fice, 1 valid, -9 not valid\*

gen vfice=.

replace vfice=1 if fice==143600 | fice==205700 | fice==205800 | fice==206100 | fice==206200 | fice==206300 | fice==206400 | ///

fice==206500 | fice==206700 | fice==206800 | fice==207100 | ///

fice==207200 | fice==207300 | fice==207400 | fice==207500 | fice==207600 | fice==207700 | ///

fice==207800 | fice==208000 | fice==208600 | fice==208700 | fice==208900 | ///

fice==209100 | fice==209200 | fice==209500 | fice==209600 | fice==209900 | fice==210200 | fice==210300 | ///

fice==210400 | fice==210500 | fice==210600 | fice==210700 | fice==210800 | fice==210900 | fice==465000 | ///

fice== 691100 | fice==795900 | fice==817500 | fice==830800 | fice==1001400 | fice==1164400 | fice== 2073900 | ///

fice==208300 | fice==2127900 | fice==2314800 | fice==2578400 | fice==3100700 | fice==3455500 | fice==90795900 | fice==99765900

mvdecode vfice, mv(-9)

tab vfice /\*valid fice, 1 valid, -9 not valid\*/

tab fice if vfice==-9 /\*list of invalid fice numbers\*/

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*valid sic, 1 valid, -9 not valid\*

gen vsic=.

replace vsic=1 if sic==110100 | sic==110200 | sic==110770 | sic==110900 | sic==111000 | ///

sic==111100| sic==111250| sic==111300 | sic==111700 | sic==111900 | sic==112100 | ///

sic==112200 | sic==112400| sic==112970 | sic==113600 | sic==115470| sic==120600 | ///

sic==121400 | sic==121800| sic==123900| sic==124200 | sic==124400| sic==124500 | ///

sic==124600 | sic==124700 | sic==124800| sic==124900 | sic==124960 | sic==124965 | ///

sic==143000 | sic==154000 | sic==260400 | sic==260410 | sic==260469 | sic==260537| ///

sic==260700 | sic==262000 | sic==262050 | sic==262230 | sic==262300| sic==262487| ///

sic==262500| sic==262570 | sic==262600| sic==262650 | sic==262700 | sic==262767 | ///

sic==263100 | sic==263124 | sic==263200| sic==263300| sic==263500 | sic==263700| ///

sic==263915 | sic==264100| sic==264150 | sic==264182| sic==264187 | sic==264300 | ///

sic==265000| sic==261200| sic==265100 | sic==265200 | sic==265300| sic==265400| ///

sic==265420| sic==265600 | sic==261575| sic==263575| sic==260500 | sic==261150| ///

sic==261450| sic==110800| sic==111500 | sic==111600 | sic==112800| sic==112900| ///

sic==112950 | sic==260250

mvdecode vsic, mv(-9)

tab vsic, missing /\*1 valid, -9 not valid\*/

tab sic if vsic==-9 /\*list of invalid sic numbers\*/

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*valid collection term, 1 is valid, 2 is not valid\*

\*change term as needed\*

gen vct = .

\*select one of the following\*

replace vct = 1 if coltm ==1

replace vct = 2 if coltm !=1

tab vct /\*1 is valid, 2 is not valid\*/

tab fice if vct==2 /\*1 is valid, 2 is not valid\*/

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*valid collection year, 1 is valid, 2 is not valid\*

\*change year as needed\*

gen vcy = .

\*select one of the following\*

replace vcy = 1 if colyr == 2018

replace vcy = 2 if colyr != 2018

tab vcy /\*1 is valid, 2 is not valid\*/

tab fice if vcy==2 /\*1 is valid, 2 is not valid\*/

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*valid ID Type, valid 1, invalid 2\*

gen vtype=.

replace vtype = 1 if idt ==1 | idt==2| idt==3

replace vtype = 2 if idt !=1 & idt !=2 & idt!=3 & idt !=.

tab vtype /\*1 is valid, 2 not valid\*/

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*valid SSN, 1 is valid, 2 is 8 digits, 3 is 10 or more digits, 4 is 7 or less digits\*

\*8 digits may indicate missing leading 0\*

/\*SSA will not issue SSNs beginning with the number “9”. SSA will not issue SSNs

beginning with the number “666” in positions 1 – 3. SSA will not issue SSNs beginning

with the number “000” in positions 1 – 3. ... SSA will not issue SSNs with the

number “0000” in positions 6 – 9.\*/

gen vssn =idn

tostring vssn, replace

\*testing for lenght of string\*

\*note, on import, the string is destrung, which will drop leading 0s\*

\*when the string is re-instated the 0s remain dropped\*

\*this helps identify IDNs with too many leading zeros but an IDT of SSN\*

gen vssn2=.

replace vssn2=1 if strlen(vssn)==9

replace vssn2=2 if strlen(vssn)==8

replace vssn2=3 if strlen(vssn)>=10

replace vssn2=4 if strlen(vssn)<=7

tab vssn2

\*validate against id type\*

tab vssn2 if idt==1 &(vssn2==3|vssn2==4) /\*number of invalid IDNs with ID Type =SSN\*/

tab fice if idt==1 &(vssn2==3|vssn2==4) /\*fice with number of invalid IDNs with ID Type =SSN\*/

tab vssn2 idt

\*validate against id type\*

tab vssn2 if idt==1 &(vssn2==3|vssn2==4) /\*number of invalid IDNs with ID Type =SSN\*/

tab fice if idt==1 &(vssn2==3|vssn2==4) /\*fice with number of invalid IDNs with ID Type =SSN\*/

tab vssn2 idt

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*validate TSSN\*

\*tssn should be 10 characters\*

gen vtssn=.

replace vtssn=3 if strlen(tssn)>10

replace vtssn=2 if strlen(tssn)<10

replace vtssn=1 if strlen(tssn)==10

tab vtssn /\*1 is valid, 2 & 3 not valid\*/

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*validate TSSN2\*

\*tssn2 should be 7 characters\*

gen vtssn2=.

replace vtssn2=3 if strlen(tssn2)>7

replace vtssn2=2 if strlen(tssn2)<7

replace vtssn2=1 if strlen(tssn2)==7

tab vtssn2 /\*1 is valid, 2 & 3 not valid\*/

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*valid campusID 1 valid, 2 not valid\*

\*campus id is 10 or less characters\*

gen vcamp=.

replace vcamp=2 if strlen(tssn)>=11

replace vcamp=1 if strlen(tssn)<=10

tab vcamp /\*1 is valid, 2 not valid\*/

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*valid SASID, code missing values as 9\*

gen vsasid =sasid

mvencode vsasid, mv(9)

tostring vsasid, replace

\*valid SASID\*, 1 and 2 are valid, 3 and 4 are not valid

gen vsasid2=.

replace vsasid2=4 if strlen(vsasid)>10 /\* not valid, too long\*/

replace vsasid2=3 if strlen(vsasid)<10 &strlen(vsasid)>1 /\* not valid, too short\*/

replace vsasid2=2 if strlen(vsasid)==10 /\*valid\*/

replace vsasid2=1 if strlen(vsasid)==1 /\*valid coded as 9 missing value\*/

tab vsasid2

tab fice if vsasid2>=3 /\*fice with SASID not valid\*/

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*valid EOTS flag 1 valid, 2 not valid\*

gen vmatchflag=.

replace vmatchflag=1 if (eotsflg==1|eotsflg==2|eotsflg==3) /\*valid\*/

replace vmatchflag=2 if (eotsflg<1|eotsflg>3|eotsflg==.) /\* not valid\*/

tab vmatchflag

tab eotsflg

tab fice if vmatchflag==2 /\*fice with not valid\*/

/\*EOTS-EIS Match Flag

1= Record contains EOTS data and is matched to a record in EIS

2= Record contains EOTS data and is NOT matched to a record in EIS

3= Record is a placeholder that contains no EOTS data and is matched to a record in EIS

The EOTS-EIS Match Flag is used to identify enrollment patterns as well as to assist in managing and validating file submissions.

For most record the flag will be set to 1. For an EOTS record without and matching EIS record the flag should be set to 2. When set

to 3 only student identification data is reported all other fields are reported as 0 or blank as appropriate.\*/

\*term academic standing for privates only\*

/\*blank is not necessarily invalid, but a high rate of blanks for any one

college could be a data error\*/

/\*1 is valid, 2 & 3 are not valid\*/

gen vtas=.

replace vtas=1 if sector==2 & tacadst=="G"

replace vtas=1 if sector==2 & tacadst=="P"

replace vtas=1 if sector==2 & tacadst=="N"

replace vtas=1 if sector==1 & tacadst==""

replace vtas=3 if sector==1 & tacadst!=""

replace vtas=2 if sector==2 & tacadst==""

tab vtas

tab fice if vtas==2 & sector==2 /\*2 is not valid = wrong value at private\*/

tab fice if vtas==3 & sector==1 /\*3 is not valid = value entered for publics\*/

tab tacadst if sector==2, missing /\*any letters/symbols other than G, P and N are invalid- errors\*/

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

/\*blank is not necessarily invalid, but a high rate of blanks for any one

college could be a data error\*/

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*first math or english grade\*

\*A, B, C, D, F and P are valid\*

\*blank is valid\*

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

gen vmath=.

replace vmath=1 if fstmathg=="A"

replace vmath=1 if fstmathg=="B"

replace vmath=1 if fstmathg=="C"

replace vmath=1 if fstmathg=="D"

replace vmath=1 if fstmathg=="F"

replace vmath=1 if fstmathg=="P"

replace vmath=9 if fstmathg==""

tab fstmathg, missing /\*any letters other than A B C D F and P are invalid - errors\*/

tab vmath /\*1 is valid, 9 is missing\*/

gen veng=.

replace veng=1 if fstenglg=="A"

replace veng=1 if fstenglg=="B"

replace veng=1 if fstenglg=="C"

replace veng=1 if fstenglg=="D"

replace veng=1 if fstenglg=="F"

replace veng=1 if fstenglg=="P"

replace veng=9 if fstenglg==""

tab fstenglg, missing /\*any letters other than A B C D F and P are invalid- errors\*/

tab veng /\*1 is valid, 9 is missing\*\*/

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*GPA\*

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*valid CGPA. 1 valid, 2 not valid, 3 blank, 4 is 0\*

\*private not required to submit\*

/\*If the field is zero-filled, this is interpreted to mean the student has a zero

cumulative GPA (i.e. failed all courses attempted)\*/

/\*3 and 4 not invalid, but high rates at any one college could be data error\*/

gen vcgpa=.

replace vcgpa=1 if cumgpa>0.00 & cumgpa<=4.99 /\*valid\*/

replace vcgpa=3 if cumgpa==. /\*valid\*/

replace vcgpa=4 if cumgpa==0 /\*valid\*/

replace vcgpa=2 if cumgpa<0.00 | cumgpa>5.00 & cumgpa!=. /\*not valid\*/

tab vcgpa

tab vcgpa if sector==1

tab fice if vcgpa==2 & sector==1 /\*not valid\*/

tab fice if vcgpa==3 & sector==1 /\*missing\*/

tab fice if vcgpa==4 & sector==1 /\*0\*/

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*valid TGPA. 1 valid, 2 not valid, 3 blank, 4 is 0\*

/\*If the field is zero-filled, this is interpreted to mean the student has a zero

cumulative GPA (i.e. failed all courses attempted).\*/

/\*3 and 4 not invalid, but high rates at any one college could be data error\*/

gen vtgpa=.

replace vtgpa=1 if tgpa>0.00 & tgpa<=4.99 /\*valid\*/

replace vtgpa=3 if tgpa==. /\*valid\*/

replace vtgpa=4 if tgpa==0 /\*valid\*/

replace vtgpa=2 if tgpa<0.00 | tgpa>5.00 & tgpa!=. /\*not valid\*/

tab vtgpa

tab vtgpa if sector==1

tab fice if vtgpa==2 & sector==1 /\*not valid\*/

tab fice if vtgpa==3 & sector==1 /\*missing\*/

tab fice if vtgpa==4 & sector==1 /\*0\*/

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\* term credits\*

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*valid term nat credit registered\*

\*total reg credits in term\*

/\*Includes all credit hours for which the student is registered at the reporting

institution during the term. May be greater than the number of credit hours

reported in EIS. Does not include credit hours for courses which were entirely

dropped from the transcript and for which no full or partial payments were collected.

Must include courses that may not earn credit toward an award, such as audited

and remedial or developmental courses.\*/

gen vreg=.

replace vreg=1 if tcrhrrg>0 /\*valid no upper bound\*/

replace vreg=2 if tcrhrrg==0 /\*valid but high number may be a problem\*/

replace vreg=3 if tcrhrrg==. /\*not valid\*/

replace vreg=4 if tcrhrrg<0 /\*not valid\*/

tab vreg

tab fice if vreg==2 /\*valid but high number may be a problem\*/

tab fice if vreg==3 /\*not valid\*/

tab fice if vreg==4 /\*not valid\*/

\*valid term nat deg credit attempted\*

/\*Includes all credit hours for which the student is registered during the term that

can be applied to a formal award. Does not include credit hours for courses which

were entirely dropped from the transcript and for which no full or partial payments

were collected. Does NOT include courses for which degree credit cannot be earned,

such as audited and remedial or developmental courses.\*/

gen vdatt=.

replace vdatt=1 if tdcrhrat>0 /\*valid no upper bound\*/

replace vdatt=2 if tdcrhrat==0 /\*valid but high number may be a problem\*/

replace vdatt=3 if tdcrhrat==. /\*not valid\*/

replace vdatt=4 if tdcrhrat<0 /\*not valid\*/

tab vdatt

tab fice if vdatt==2 /\*valid but high number may be a problem\*/

tab fice if vdatt==3 /\*not valid\*/

tab fice if vdatt==4 /\*not valid\*/

\*valid term nat credit earned\*

/\*Include only those credit hours earned at reporting institution. Transfer credit

hours, regardless of their status, should not be included. Audited and developmental

or remedial courses should not be included.\*/

\*credits earned toward degree\*

gen vcearn=.

replace vcearn=1 if tndcrhre>0 /\*valid no upper bound\*/

replace vcearn=2 if tndcrhre==0 /\*valid but high number may be a problem\*/

replace vcearn=3 if tndcrhre==. /\*not valid\*/

replace vcearn=4 if tndcrhre<0 /\*not valid\*/

tab vcearn

tab fice if vcearn==2 /\*valid but high number may be a problem\*/

tab fice if vcearn==3 /\*not valid\*/

tab fice if vcearn==4 /\*not valid\*/

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*cum credits testing\*

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*for both cum credits, blank not valid; cum award must be >= cum earned

\*earned credits coding for blank and less than 0 (aka negative numbers)\*

\*native credits earned\*

/\*These are only those credit hours earned at reporting institution. Transfer credit

hours, regardless of their status, should not be included. Remedial courses should

not be included. \*/

\*cum credits earned toward degree in residence at the institution\*

gen vcumnatcre=.

replace vcumnatcre=1 if cncrhre>0 /\*valid no upper bound\*/

replace vcumnatcre=2 if cncrhre==0 /\*valid but high number may be a problem\*/

replace vcumnatcre=3 if cncrhre==. /\*not valid\*/

replace vcumnatcre=4 if cncrhre<0 /\*not valid\*/

tab vcumnatcre

tab fice if vcumnatcre==2 /\*valid but high number may be a problem\*/

tab fice if vcumnatcre==3 /\*not valid\*/

tab fice if vcumnatcre==4 /\*not valid\*/

\*awarded credits\*

/\*Includes all credit hours awarded by the reporting institution that can be

applied toward the requirements for a formal award. This includes native credit,

transfer credit, credit by examination, and all other forms of credit.\*/

\*cum credits from all sources toward degree\*

gen vcumnatcra=.

replace vcumnatcra=1 if ccrhra>0

replace vcumnatcra=2 if ccrhra==0

replace vcumnatcra=3 if ccrhra==.

replace vcumnatcra=4 if ccrhra<0

tab vcumnatcra

tab fice if vcumnatcra==2

tab fice if vcumnatcra==3

tab fice if vcumnatcra==4

\*logic checks\*

\*CGPA blank with credits earned\*

tab fice if cumgpa==. & cncrhre==0 & sector==1 /\*correct entry\*/

tab fice if cumgpa==. & cncrhre!=0 & sector==1 /\*incorrect entry\*/

\*CGPA=0 with credits earned\*

tab fice if cumgpa==0 & cncrhre==0 & sector==1 /\*correct entry\*/

tab fice if cumgpa==0 & cncrhre!=0 & sector==1 /\*incorrect entry\*/

\*TGPA blank with credits earned\*

tab fice if tgpa==. & tndcrhre==0 & sector==1 /\*correct entry\*/

tab fice if tgpa==. & tndcrhre!=0 & sector==1 /\*incorrect entry\*/

tab fice if tgpa==0 & tndcrhre==0 & sector==1 /\*correct entry\*/

tab fice if tgpa==0 & tndcrhre!=0 & sector==1 /\*incorrect entry\*/

\*blank CGPA with populated Term GPA\*

tab fice if cumgpa==. & tgpa!=.

\*validate total registered is equal to or more than degree credits attempted\*

/\*registered credits include degree credits in total, therefore reg credits

should be = or > than degree attempted credits\*/

gen vregatt=.

replace vregatt=1 if tcrhrrg >=tdcrhrat /\*correct\*/

replace vregatt=2 if tcrhrrg < tdcrhrat /\*incorrect\*/

tab vregatt

tab fice if vregatt==2

\*validate total registered is equal to or more than degree credits earned\*

/\*registered credits is all credits, both degree and non-degree,

therefore reg credits should be > or = degree earned credits\*/

gen veareg=.

replace veareg=1 if tcrhrrg >=tndcrhre /\*correct\*/

replace veareg=2 if tcrhrrg < tndcrhre /\*incorrect\*/

tab veareg

tab fice if veareg==2

\*validate degree credits attempted is equal to or more than degree credits earned\*

gen veaatt=.

replace veaatt=1 if tdcrhrat >=tndcrhre /\*correct\*/

replace veaatt=2 if tdcrhrat < tndcrhre /\*incorrect\*/

tab veaatt

tab fice if veaatt==2

\*testing if cum earned > term earned\*

gen vtece=.

replace vtece = 1 if cncrhre >= tndcrhre /\*correct\*/

replace vtece = 2 if cncrhre < tndcrhre /\*incorrect\*/

tab vtece

tab fice if vtece==2

\*coding that cum award > = cum earn\*

gen earnaward=.

replace earnaward=1 if ccrhra >= cncrhre /\*correct\*/

replace earnaward=2 if ccrhra < cncrhre /\*incorrect\*/

tab earnaward

tab fice if earnaward==2

\*population distribution\*

\*compare distributions to prior years to look for data changes\*

\*term GPAs\*

gen tgpadist=.

replace tgpadist=0 if tgpa==0

replace tgpadist=1 if tgpa>0 & tgpa<2

replace tgpadist=2 if tgpa>=2 & tgpa<3

replace tgpadist=3 if tgpa>=3& tgpa<4

replace tgpadist=4 if tgpa==4

replace tgpadist=5 if tgpa>4

replace tgpadist=9 if tgpa==.

tab tgpadist if sector==1

\*Cum GPAs\*

gen cgpadist=.

replace cgpadist=0 if cumgpa==0

replace cgpadist=1 if cumgpa>0 & cumgpa<2

replace cgpadist=2 if cumgpa>=2 & cumgpa<3

replace cgpadist=3 if cumgpa>=3 & cumgpa<4

replace cgpadist=4 if cumgpa==4

replace cgpadist=5 if cumgpa>4

replace cgpadist=9 if cumgpa==.

tab cgpadist if sector==1

\*grade distribution for First Math and First English\*

tab fstmathg

tab fstenglg

\*term academic standing\*

tab tacadst

tab tacadst if sector==2, missing

\*credit distributions\*

gen reg=.

replace reg=1 if tcrhrrg==.

replace reg=2 if tcrhrrg==0

replace reg=3 if (tcrhrrg>=.001&tcrhrrg<=11.99) & tcrhrrg!=.

replace reg=4 if (tcrhrrg>=12&tcrhrrg<=16.99) & tcrhrrg!=.

replace reg=5 if tcrhrrg>=17 & tcrhrrg!=.

replace reg=6 if tcrhrrg<0

tab reg

gen attp=.

replace attp=1 if tdcrhrat==.

replace attp=2 if tdcrhrat==0

replace attp=3 if (tdcrhrat>=.001&tdcrhrat<=11.99)

replace attp=4 if (tdcrhrat>=12&tdcrhrat<=16.99)

replace attp=5 if tdcrhrat>=17

replace attp=6 if tdcrhrat<0

tab attp

gen earn=.

replace earn=1 if tndcrhre==.

replace earn=2 if tndcrhre==0

replace earn=3 if (tndcrhre>=.001&tndcrhre<=11.99)

replace earn=4 if (tndcrhre>=12&tndcrhre<=16.99)

replace earn=5 if tndcrhre>=17

replace earn=6 if tndcrhre<0

tab earn

gen cearn=.

replace cearn=1 if cncrhre==.

replace cearn=2 if cncrhre==0

replace cearn=3 if (cncrhre>=.001&cncrhre<=30.99)

replace cearn=4 if (cncrhre>=31&cncrhre<=60.99)

replace cearn=5 if (cncrhre>=61&cncrhre<=90.99)

replace cearn=6 if (cncrhre>=91&cncrhre<=119.99)

replace cearn=7 if cncrhre>=120

replace cearn=8 if cncrhre<0

tab cearn

gen award=.

replace award=1 if ccrhra==.

replace award=2 if ccrhra==0

replace award=3 if (ccrhra>=.001&ccrhra<=30.99)

replace award=4 if (ccrhra>=31&ccrhra<=60.99)

replace award=5 if (ccrhra>=61&ccrhra<=90.99)

replace award=6 if (ccrhra>=91&ccrhra<=119.99)

replace award=7 if ccrhra>=120

replace award=8 if ccrhra<0

tab award

\*EOTS flag\*

tab eotsflg

\*prep EOTS for merge with EIS\*

drop sic subcamp tssn tssn2 idt campusid sasid tcrhrrg tdcrhrat tndcrhre tgpa cncrhre cumgpa ccrhra

drop tacadst fstmathg fstenglg source\_file sector-earnaward

tab fice

tab fice eotsflg

\*prep EIS for merge with EOTS\*

drop subcamp tssn idt campusid sasid sasid-ptax decrse deloc cbsc hsgyr hsgpa-sending\_opeid milstat-vtlal

tab fice

merge m:m idn fice using "M:\\\EIS\_xxxx\_xxx\_merge.dta", nolabel

label define merge ///

1 "EOTS only" 2 "EIS only" 3 "EIS and EOTS"

label values \_merge merge

\*eots flag analysis post merge\*

/\*1= Record contains EOTS data and is matched to a record in EIS

2= Record contains EOTS data and is NOT matched to a record in EIS

3= Record is a placeholder that contains no EOTS data and is matched to a record in EIS\*/

tab fice eotsflg

tab fice eotsflg if \_merge==3 & eotsflg==2 /\*matched record with EOTS flag of 2 is error as EOTS record was in EIS when it is coded as not being in EIS\*/

tab fice eotsflg if \_merge==1 & eotsflg==1/\*unmatched record with EOTS flag of 1 is error, record was coded as in EIS but record was not found in EIS\*/