**TRAC(T) FTE calculation – algorithm to estimate completed FTE for modules**

**with unknown outcome from the HESA student record**

**Step 1: Carry out the data extraction as outlined in paragraphs 1 to 11 in Annex N of the HE Data Requirements Circular (latest version available** [**here**](https://www.hefcw.ac.uk/en/statistics-and-data/hesa-data-guidance/)**).**

**Step 2: Create dataset to calculate proportions in each module outcome**

**category**

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

Retain data for all module outcomes except those with unknown outcome (those with

STUDENTONMODULE.MODSTAT=1, 2 will be in the resulting dataset). Make adjustment for year of course greater than 6.

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

if modout = '6' then delete

if yearprg ≥ 6 then yearprg = 6

**Step 3: Summarise the number of observations in each module outcome**

**category**

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

Create variables to store the FTE in each module outcome category.

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

modout1=0

modout2=0

modout3=0

modout4=0

if modout='1' then modout1\_fte=xfte01

else if modout='2' then modout2=xfte01

else if modout='3' then modout3=xfte01

else if modout='4' then modout4=xfte01

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

The results for STUDENTONMODULE.MODSTAT=1 (continuing from previous year) are to be applied to cases where STUDENTONMODULE.MODSTAT=3 (continuing into next year), all of which have unknown module outcomes. Therefore, recode STUDENTONMODULE.MODSTAT=1 to STUDENTONMODULE.MODSTAT=3 so that results can be merged back onto extracted data to the rows of data the proportions are to be applied to.

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

if modstat=’1’ then modstat=’3’

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

Sum each variable over institution, mode, level, ASC, module subject (HECoS code), year of course and module status. A new variable FREQ containing the

number of observations in each institution, mode, level, ASC, module subject, year

of course and module status combination is also calculated in this step.

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

summary by instid hesmode level asc modsbj yearprg modstat

variables modout1 modout2 modout3 modout4 xfte01

output dataset=outfte1

frequency count=FREQ

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

Do the same by institution, mode, level, ASC, year of course and module status; by

institution, mode, level, ASC and module status; by institution, mode, level and

module status; and by institution, mode and level.

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

summary by instid hesmode level asc yearprg modstat

variables modout1 modout2 modout3 modout4 xfte01

output dataset=outfte2

frequency count=FREQ

summary by instid hesmode level asc modstat

variables modout1 modout2 modout3 modout4 xfte01

output dataset=outfte3

frequency count=FREQ

summary by instid hesmode level modstat

variables modout1 modout2 modout3 modout4 xfte01

output dataset=outfte4

frequency count=FREQ

summary by instid hesmode level

variables modout1 modout2 modout3 modout4 xfte01

output dataset=outfte5

frequency count=FREQ

**Step 4: Calculate proportions in each module outcome category, rounded to 2**

**decimal places, for each of the five summary datasets.**

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

The proportions by institution, mode, level, ASC, year of course and module status;

institution, mode, level, ASC and module status; institution, mode, level and module

status; or institution, mode and level will be used to fill in the gaps where proportions

are missing. This will be where a particular module subject, year of course, ASC or

module status has no data to use in calculating proportions or there are less than 20

data rows available to use in calculating proportions.

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

data outfte1

if FREQ≥20 then do

p\_modout1=round(modout1/xfte01,0.01)

p\_modout2=round(modout2/xfte01,0.01)

p\_modout3=round(modout3/xfte01,0.01)

p\_modout4=round(modout4/xfte01,0.01)

end

data outfte2

if FREQ≥20 then do

p\_modout1\_yp=round(modout1/xfte01,0.01)

p\_modout2\_yp=round(modout2/xfte01,0.01)

p\_modout3\_yp=round(modout3/xfte01,0.01)

p\_modout4\_yp=round(modout4/xfte01,0.01)

end

data outfte3

if FREQ≥20 then do

p\_modout1\_asc=round(modout1/xfte01,0.01)

p\_modout2\_asc=round(modout2/xfte01,0.01)

p\_modout3\_asc=round(modout3/xfte01,0.01)

p\_modout4\_asc=round(modout4/xfte01,0.01)

end

data outfte4

p\_modout1\_modstat=round(modout1/xfte01,0.01)

p\_modout2\_modstat=round(modout2/xfte01,0.01)

p\_modout3\_modstat=round(modout3/xfte01,0.01)

p\_modout4\_modstat=round(modout4/xfte01,0.01)

end

data outfte5

p\_modout1\_modlev=round(modout1/xfte01,0.01)

p\_modout2\_modlev=round(modout2/xfte01,0.01)

p\_modout3\_modlev=round(modout3/xfte01,0.01)

p\_modout4\_modlev=round(modout4/xfte01,0.01)

end

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

Merge these five datasets containing proportions together and where the proportions

by module subject are missing, replace with the proportions by mode, level, ASC,

year of course and module status.

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

merge outfte1 outfte2 outfte3 outfte4 outfte5

if p\_modout1=**.** then do

p\_modout1=p\_modout1\_yp

p\_modout2=p\_modout2\_yp

p\_modout3=p\_modout3\_yp

p\_modout4=p\_modout4\_yp

end

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

Where the proportions by mode, level, ASC, year of course and module status are

also missing, replace with the proportion by mode, level, ASC and module status.

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

if p\_modout1=**.** then do

p\_modout1=p\_modout1\_asc

p\_modout2=p\_modout2\_asc

p\_modout3=p\_modout3\_asc

p\_modout4=p\_modout4\_asc

end

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

Where the proportions by mode, level, ASC and module status are also missing,

replace with the proportion by mode, level and module status.

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

if p\_modout1=**.** then do

p\_modout1=p\_modout1\_modstat

p\_modout2=p\_modout2\_modstat

p\_modout3=p\_modout3\_modstat

p\_modout4=p\_modout4\_modstat

end

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

Where the proportions by mode, level and module status are also missing, replace

with the proportion by mode and level.

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

if p\_modout1=**.** then do

p\_modout1=p\_modout1\_modlev

p\_modout2=p\_modout2\_modlev

p\_modout3=p\_modout3\_modlev

p\_modout4=p\_modout4\_modlev

end

**Step 5: These proportions are merged back onto the original extracted dataset in and used to calculate the estimated amount of FTE in each module outcome category for rows of data with STUDENTONMODULE.MODOUT = 6**