

HE entry - Evaluating the predictive power and variables of the HEAT Groups model

HEAT uses its own classification system - **The HEAT Groups** - for grouping outreach participants to reflect their likelihood of enrolling in Higher Education.

The HEAT Groups use a basket of measures which indicate socio-economic deprivation and prior attainment level.*

*For a full HEAT Groups methodology please contact the HEAT service. Members can download a copy from the HEAT File Store

Attainment

Disadvantage

	High	Low
Low	Group 1	Group 3
Medium	Group 2b	Group 4b
Highest	Group 2a	Group 4a

HEAT commissioned the **Behavioural Insights Team (BIT)*** to evaluate the predictive power of the HEAT Groups and the variables that contribute to the classification. * <https://www.bi.team>

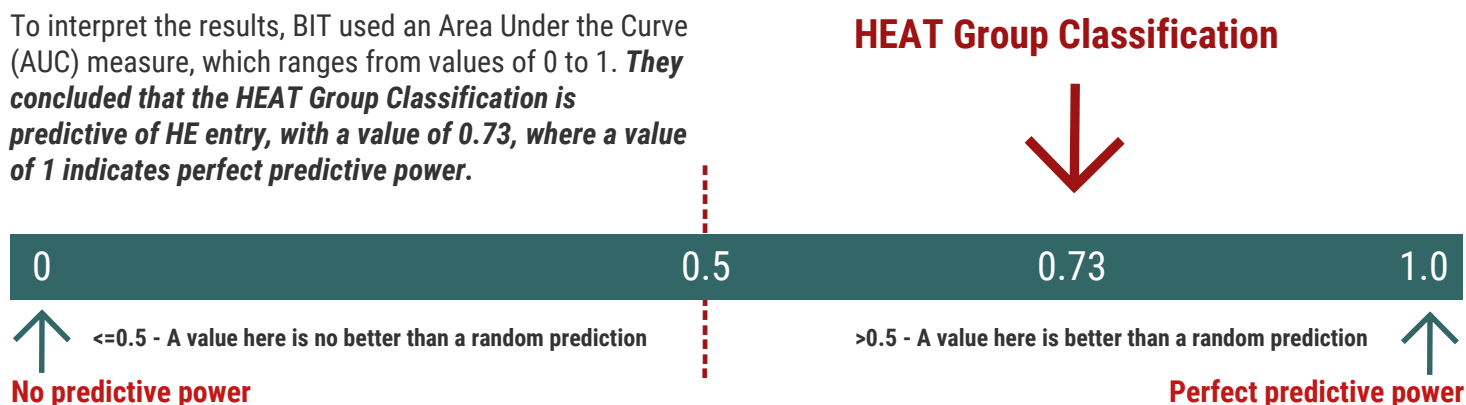
Findings

- *Members can be confident that the HEAT Groups Classification is predictive of HE entry*
- *It continues to be important for the HEAT membership to work with the Department for Education to continue to gain access to data on prior attainment*
- *Although there has been debate across the membership about the collection and wording of a 'First Generation HE' self-declaration by outreach participants, this research points to its predictive importance. Members may wish to consider collecting this information if they do not already do so*
- *This research points to the importance of Low School HE Participation as a predictor of HE entry. This information is available in the Schools Planning Dataset available to all members. Members may wish to revisit this as a targeting criteria. In addition the Research Group and the HEAT data team will consider this indicator for future thematic research*

Predictive Value

Using a machine learning algorithm, BIT constructed a decision tree to calculate prediction errors in the HEAT Groups.

To interpret the results, BIT used an Area Under the Curve (AUC) measure, which ranges from values of 0 to 1. ***They concluded that the HEAT Group Classification is predictive of HE entry, with a value of 0.73, where a value of 1 indicates perfect predictive power.***



BIT performed a cross-check on this machine learning value of 0.73 by taking the top 10% of outreach participants most likely enter to HE (as defined by the tree-structure model they constructed) and looked at their actual HE entry rate. If the HEAT classifications were to have perfect predictive power this group would enter HE at a rate of 100%. In actual fact 82% of that sub-group went on to enter HE. However, since the overall entry rate for the whole cohort was 53%, the HEAT Groups classification was clearly better able to predict HE entry for that top 10% sub-group than just assigning those individuals to a random prediction.

Although the HEAT Group Classification does not have 'perfect' predictive power, it should be remembered that HEAT Groups cannot control for unknowable, and currently unmeasurable, factors affecting the decision making of individuals such as students' personal motivation or input from family, friends and teachers. These are likely to account for the 0.23 missing in the machine-learning value and for the 18% of the cross-check group who did not enrol in HE. Indeed the findings in the table below, about the relative importance of the measures that make up the HEAT Group model, suggest that it is factors at an individual level which have most predictive power.

Relative importance of individual measures within the HEAT Groups Classification

This table shows the relative importance of each of the ten HEAT group components for the groups' overall predictive power. The values in the second column sum to 1, so that we can directly interpret the values as relative importance.

A low value does not indicate that a variable makes no contribution to the measure, only that it has less relative importance in the context of this research. We know for instance that the gap between Free School Meals (FSM) pupils and non-FSM gaining 5 good GCSEs continues to be significant for targeting individuals.

In the context of this research, it should be noted that the top 4 variables account for 91% of the predictive power of the HEAT Groups classification system.

It should also be noted that individual-level measures, such as prior attainment and self-declaration by the student that they will be first in family to enter HE, are more predictive of HE entry than aggregate measures such as IMD.

Variable	Importance
High GCSE*	0.5
GCSE score below 352	0.28
School has low HE Participation rate	0.09
First Generation HE**	0.04
Free School Meals	0.02
POLAR4 quintiles 1&2	0.02
POLAR3 quintiles 1&2	0.02
IDACI Index quintiles 1&2	0.01
IMD Index quintiles 1&2	0.01
ACORN groups M, O, P or Q**	0.01

* Achieved 5 GCSEs including English and Maths at grades 4-9 (formerly A*-C)

**self-declared by HEAT outreach participants

** For an explanation of ACORN and its groups. please see <https://acorn.caci.co.uk/>