ARTS TUTORING PROGRAMME:

2024/2025 IMPACT EVALUATION REPORT

Attainment Raising Programmes

Network for East Anglian Collaborative Outreach (neaco)

Project description

Arts Tutoring is a year-long programme consisting of weekly or biweekly 1- or 2-hour sessions, designed for Year 10 and 11 students. Participants were selected by their schools based on their receipt of Free School Meals (FSM) and/or being from underrepresented groups (URG). The programme was delivered across 5 schools in East Anglia (Norfolk), providing one-to-one support during school time to a total of 47 participants. The programme was delivered in partnership with our (arts) Higher Education Champion based in schools in the East of England.

The main aim of the programme is to raise students' attainment levels in creative subjects by helping them develop research skills they can apply across the curriculum and, particularly, within creative disciplines. The programme also seeks to increase student's motivation to pursue the subject at A level and to enhance their understanding of creative higher education (HE) routes.

The programme does not have a fixed number of sessions, as these depend on the needs and availability of the students and schools. Sessions take the form of one-to-one tutoring, during which students focus on their art subject work and can ask questions to the delivering Higher Education Champion (HEC) throughout.

Evaluation approach

The programme was underpinned by a Theory of Change. All activity was logged on the Higher Education Access Tracker (HEAT) and made use of the HEAT Attainment Raising Typology to code activity. The evaluation focused on a pre-and-post design, complemented by a comparison of predicted GCSE grades in arts subjects collected before the intervention with predicted or actual grades (for Year 11 students) obtained after the intervention.

The pre-and-post surveys looked at student cognitive, metacognitive and arts skills (and how these affected the learners' confidence) as well as at academic self-efficacy. Survey questions used were based on TASO's Access and Success Questionnaire (ASQ). Additionally, some open-ended qualitative questions were included to capture the learners' main takeaways from the project, allowing them to reflect on their experiences more freely. The evaluation tracked the changes in these specific skills and outcomes before and after the intervention, and collected information on the learners' perceived impact of the project.

Pre- and post-project surveys were sent to 47 Year 10 and Year 11 students across five schools of Norfolk, East Anglia (see Participants section) before and after their

participation in the Arts Tutoring programme. Each school had between 1 and 8 participating students, all of whom received one-to-one tutoring through the programme. Surveys were available in both electronic and paper formats, with a preference for paper, which helped mitigate issues related to technology access in the classroom and supported a higher response rate. Grades were gathered by the relevant HEC, based on information supplied by school staff once GCSE were available at the end of the programme. This amounts to an OfS Standards of Evidence Type 2 approach that generates empirical evidence but cannot provide an insight into the specific causal impact of the project.

To analyse impact, paired Wilcoxon tests were conducted to compare pre- and post-survey responses, as well as predicted versus actual grades. In addition, a linear mixed-effects model (LMM) was run on the grades data to explore whether the number of contact hours was associated with greater grade improvement over time. The sample size of matched responses (see section below) is sufficient to detect moderate to large changes, though smaller effects may not reach statistical significance. Therefore, the findings provide useful insights into the students who participated, while generalisations beyond this group should be made carefully.

Results

Participants

The programme was delivered to 47 students, of which 14 were Year 10 learners and 33 were Year 11 learners. Out of these, 44 completed the pre-programme survey (93.6% response rate) and 32 completed the post-programme survey (68.1% response rate). In total, 30 students completed both the pre- and post- surveys, accounting for a 63.8% overall response rate. Of these, 6 were from Year 10 students and 24 were from Year 11 students. The small year-group sizes prevented separate analyses; therefore, responses were combined for an overall impact analysis.

Findings and discussion

This section first presents the survey results followed by an analysis of the GCSE grade data. It should be noted that the two datasets are not directly comparable, as the grade information was available only for the Year 11 group. Moreover, some Year 11 students may not have completed both, or any, of the survey assessments, but their grades were still collected. Given the small sample sizes, and to maximise the use of available data, all grades collected were included in the grade comparison analysis, rather than restricting it to those who also have matched survey responses.

Starting with the survey data, the figures below, constructed from the 30 matched pre- and post- survey data, illustrate several key findings of the programme:

KEY FINDING 1: Some positive changes were found in the students' self-perceived cognitive, metacognitive and arts-related skills, as well as in their self-efficacy (both post-16 and HE), after participating in the Arts Tutoring programme, although none reached statistical significance.

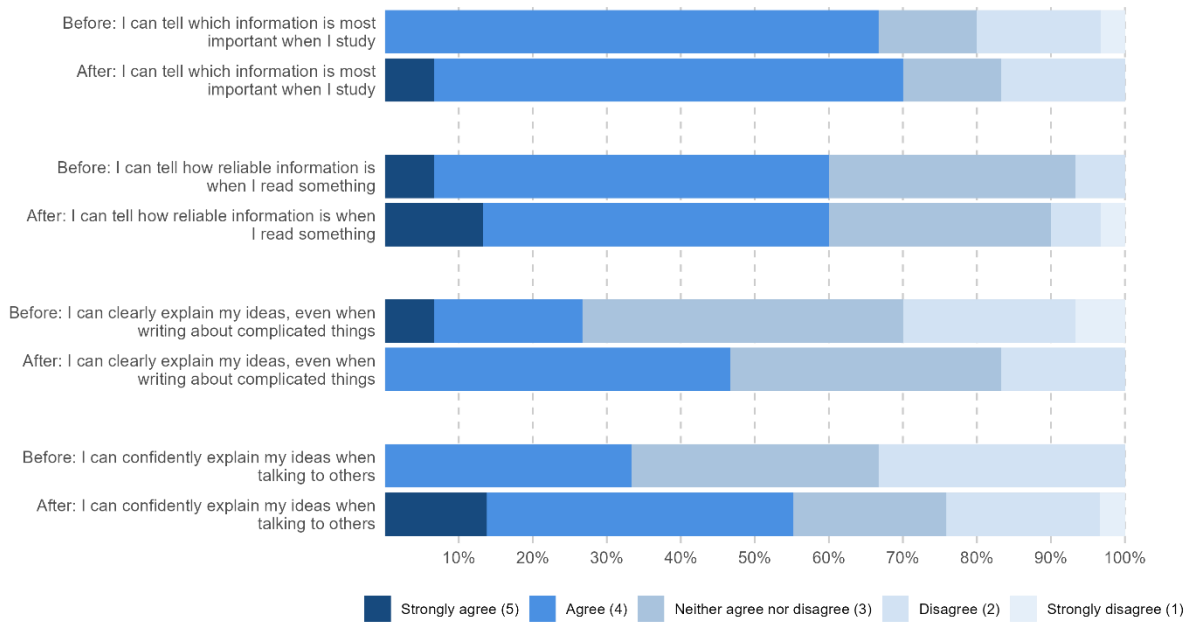

Figure 1: Cognitive skills. Wilcoxon signed-rank tests revealed no significant differences between the pre- and post-survey results for any of the cognitive skills questions (*p = 0.396, p = 0.948, p = 0.08 and p = 0.110,* respectively). No significance was found when questions were combined and treated as a separate data point for the overall 'Cognitive skills' category (*p = 0.06)*.

While the statistical tests conducted at the block level indicated a non-significant positive change in the students' self-reported perspectives on their cognitive skills, the result approached statistical significance (*p = 0.06*). Further insight is provided by the question-level analysis, which revealed that this overall effect was primarily driven by the third question (i.e., on explaining their ideas clearly), which was the only item within the block also close to showing a statistically significant improvement (*p = 0.08*). This suggests that the programme might have contributed to improvements in students' communication and articulation of ideas, though the small sample size may have limited the ability to detect statistically significant effects. A larger sample size would help clarify whether the positive trends observed represent a genuine programme effect or are due to sampling variation. However, given the nature of the Arts Tutoring programme, which by design involves relatively small cohorts, it might not always be feasible to achieve large

participant numbers. Future iterations could therefore consider other small-n methodologies to shed further light on these results and capture more nuances of the potential impact (see Recommendations section).

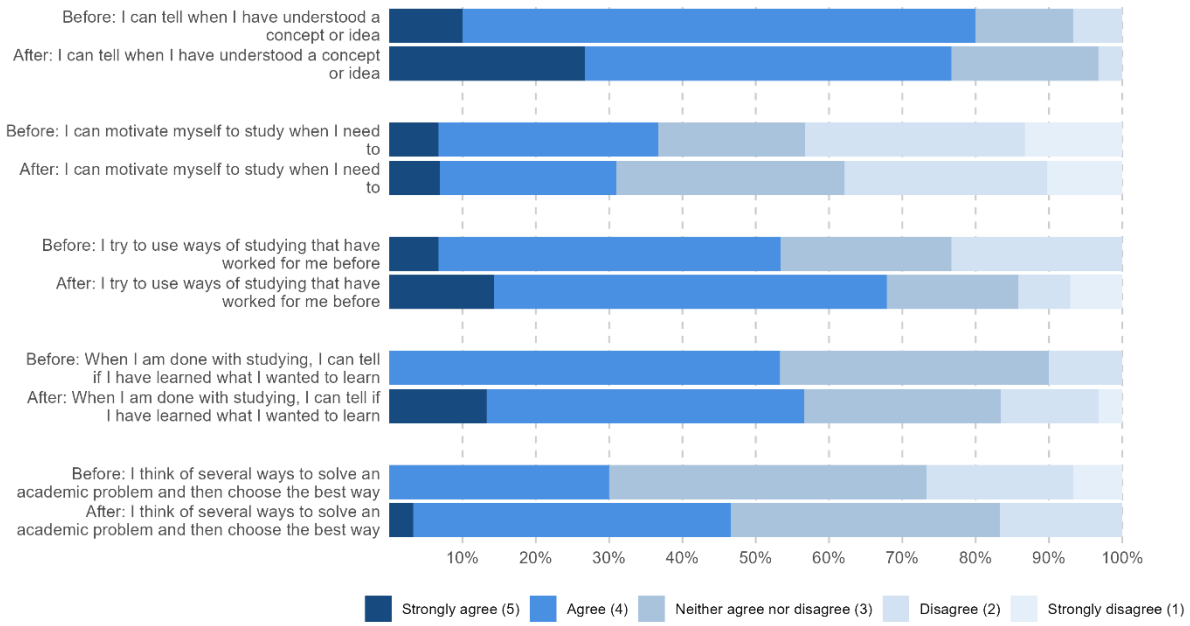

Figure 2: Metacognitive skills. Wilcoxon signed-rank tests revealed no significant differences between the pre- and post-survey results of the 'Metacognitive skills' block (*p = 0.07*). Regarding question-level analyses, a significant positive difference was found for the fifth question (*p = 0.02*), while no significant differences were observed in the rest of the questions in this block (*p = 0.263. p = 0.683, p = 0.336 and p = 0.660*, respectively).

Similar to the effect observed in the Cognitive Skills block, the block-level analysis of metacognitive skills also revealed a result close to statistical significance (*p = 0.07*). A more fine-grained, question-level analysis showed that this effect was also mainly driven by one question, in this case, the fifth question (i.e., relating to thinking and solving academic problems), which did reach statistical significance (*p = 0.02*). However, as with the previous block, the small sample size limits the strength of any conclusions that can be drawn. It therefore remains unclear whether this reflects a genuine programme effect, whether additional effects may not have been captured, or whether the observed results are due to sampling variation. Accordingly, these findings should also be interpreted with caution.

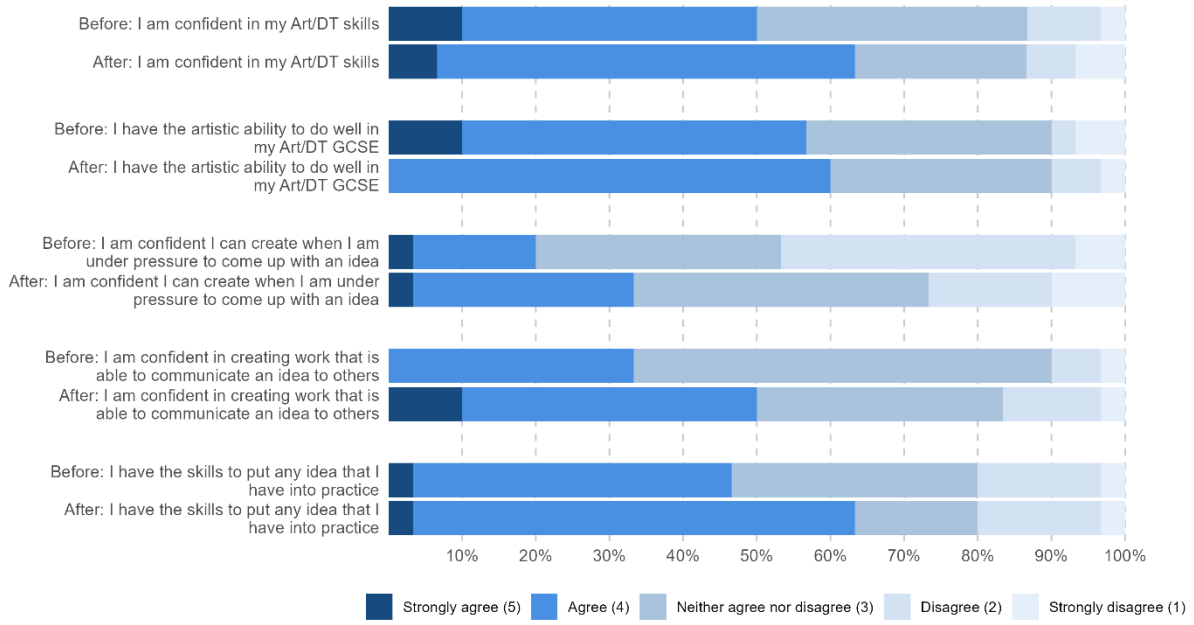

Figure 3: Arts skills. Wilcoxon signed-rank tests revealed no significant differences between the pre- and post-survey results for any of the arts skills questions (*p = 0.660, p = 0.660, p = 0.246, p = 0.403* and *p = 0.282*, respectively). No significance was found when questions were combined and treated as a separate data point for the overall 'Arts skills' category (*p = 0.392)*.

In the 'Arts skills' block, some positive shifts were observed across all questions, although none reached statistical significance. Improvements were particularly seen in questions 1, 3, and 5, all relating to students' confidence and practice of arts skills, which, as discussed below in the qualitative feedback sections, are areas that students highlighted themselves when asked openly about the programme's impact. In contrast, less change was observed in question 2, which concerns perceived artistic ability, and question 4, which relates to communicating ideas to others. The limited movement in question 2 aligns with students' broader perceptions of their academic ability, as illustrated in Figures 4 and 5 below. For question 4, smaller changes likely reflect the programme's lack of focus on communication beyond student–tutor interactions, such as presenting to peers or wider audiences, which constrains perceived progress in this domain.

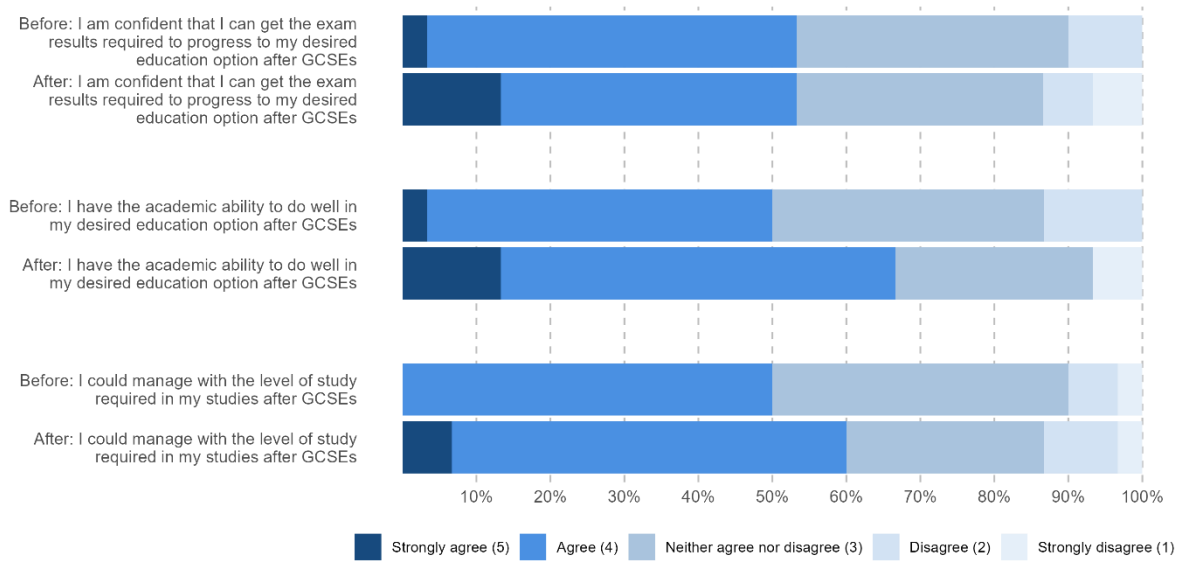

Figure 4: Self-efficacy (post-16). Wilcoxon signed-rank tests revealed no significant differences between the pre- and post-survey results for any of the self-efficacy (post-16) questions ($p = 1$, $p = 0.165$ and $p = 0.376$, respectively). No significance was found when questions were combined and treated as a separate data point for the overall 'Self-efficacy (post-16)' category ($p = 0.310$).

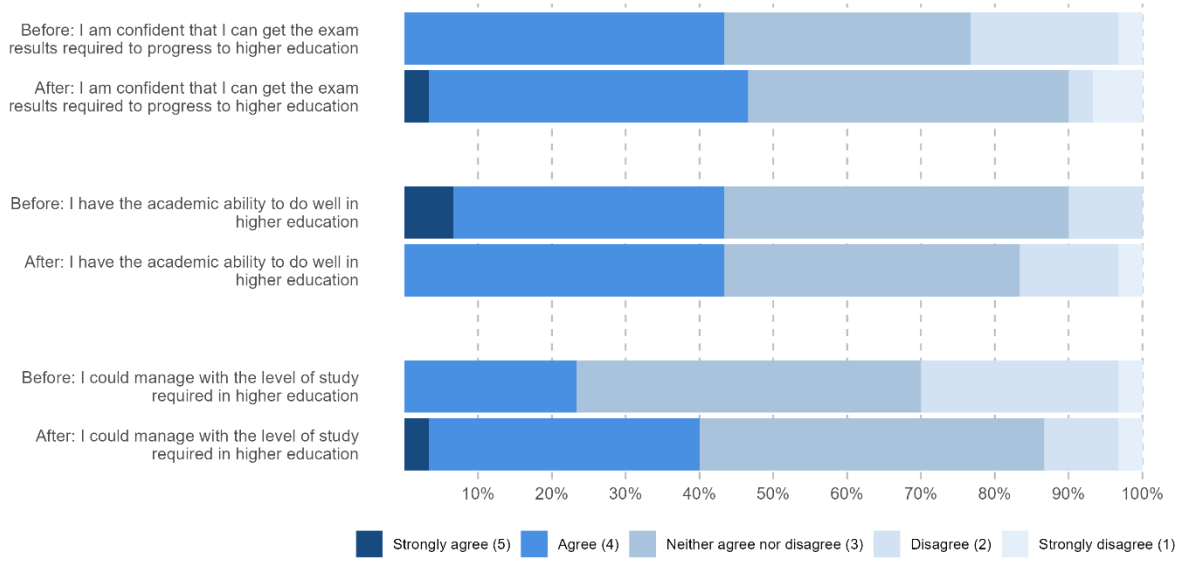

Figure 5: Self-efficacy (HE). Wilcoxon signed-rank tests revealed no significant differences between the pre- and post-survey results for any of the self-efficacy (HE) questions ($p = 0.239$, $p = 0.474$ and $p = 0.06$, respectively). No significance was found when questions were combined and treated as a separate data point for the overall 'Self-efficacy (post-16)' category ($p = 0.453$).

With regards to self-efficacy, some positive shifts can be observed both at post-16 and HE levels, although, again, none reached the threshold for statistical significance.

Nonetheless, some interesting observations can be drawn from these two blocks. Firstly, the proportion of 'strongly agree' and 'agree' responses was higher in the post-16 block than in the HE block, which aligns with the expected progression in students' academic confidence, that is, greater confidence in coping with nearer academic challenges (post-16) than with those perceived as more distant or demanding (HE). Interestingly, in both blocks, a considerable increase in positive responses was observed in the third question (i.e., relating to coping with academic demands), which in turn is aligned with the rise in confidence reported in the qualitative responses (to be discussed later) and the positive changes observed across the skills blocks. Another notable finding, consistent with the perceived impact data (see Figure 6 below), is that close to 70% of students agreed or strongly agreed that they have the academic ability to do well in post-16 education, whereas this proportion drops to just over 40% for HE. This pattern mirrors the perceived impact results, where only 43.3% of students agreed or strongly agreed that the programme had increased their higher education aspirations. This disparity may reflect students' lower confidence in their ability to succeed at the HE level, or perhaps a lack of information, advice, and guidance (IAG) regarding HE options, an aspect that future iterations of the programme could explore and address further (see Recommendations section).

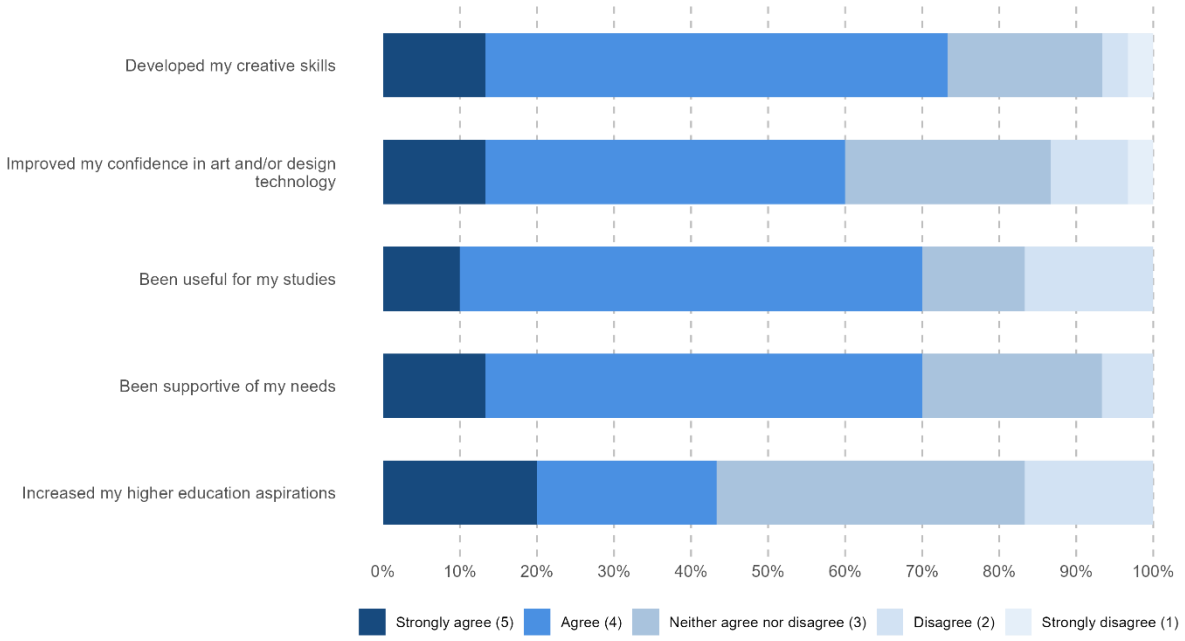

Figure 6: Perceived impact of the Arts Tutoring programme.

In addition to the impact on HE aspirations, Figure 6 offers further insights into students' perceptions of the Arts Tutoring programme's impact. When asked directly whether the programme had helped them develop their creative skills, 73.3% of respondents agreed or strongly agreed with this claim, followed closely by a 70% of positive responses indicating that it had been useful and supportive for their studies. These findings align

closely with the themes emerging from the open-ended qualitative responses (see Table 1), where students highlighted the improvement of skills applied to and/or related to arts (50%), as well as the benefits of having external support (12.5%) and a dedicated work-focused space (12.5%). This convergence between the different sets of questions suggests that the programme fostered both creative skill development and a supportive learning environment. A slightly lower, yet still substantial, 60% of students agreed or strongly agreed that the programme had improved their confidence in art/or design technology. Once again, this matches the qualitative findings, where increased confidence emerged as the third most frequently mentioned takeaway from the programme, mentioned by 25% of respondents in the open-ended questions section.

Biggest takeaway of the programme	Percentage of responses[1]
Improved skills applied to/related to arts	50%
Increased creativity	31.5%
Increased confidence	25%
Having external help	12.5%
Having a work-focused space	12.5%

Table 1. Summary of topics raised in the intervention learners' responses to open-ended questions.

Lastly, in addition to the feedback responses discussed, several students provided brief testimonials reflecting on their participation in the programme. A selection of these, from students across different participating schools, is presented below:

"I found the programme really help me with my confidence in art and exploration." – Student at Hewett Academy

"[The biggest takeaway from the programme was] exploring new ideas in my artwork." – Student at Wayland Academy

"This should happen in other lessons and subjects." – Student at Hewett Academy

[1] Please note 16 out of 30 students decided to answer the open-ended questions.

Moving now to the grades data, this analysis focuses on the Year 11 cohort, as predicted and actual GCSE grades were only available for this group at the time of writing. Grade data were obtained for 27 of the 33 participating Year 11 students, as some results could not be collected or the students did not ultimately sit the exams. To reiterate, the grade data do not correspond directly to the matched survey responses, as not all Year 11 students completed both surveys. Therefore, to maximise the use of available data given the small sample size, all grades collected were included in the analysis.

Students' grade data were examined to assess the programme's potential academic performance impact. Specifically, the analysis explored whether participation in the programme was associated with improvements between predicted, pre-intervention grades and actual GCSE, post-intervention grades. The results from this analysis lead to a second key finding of the programme:

KEY FINDING 2: There was a slight overall improvement in post-intervention grades compared to pre-intervention, although this change did not reach statistical significance.

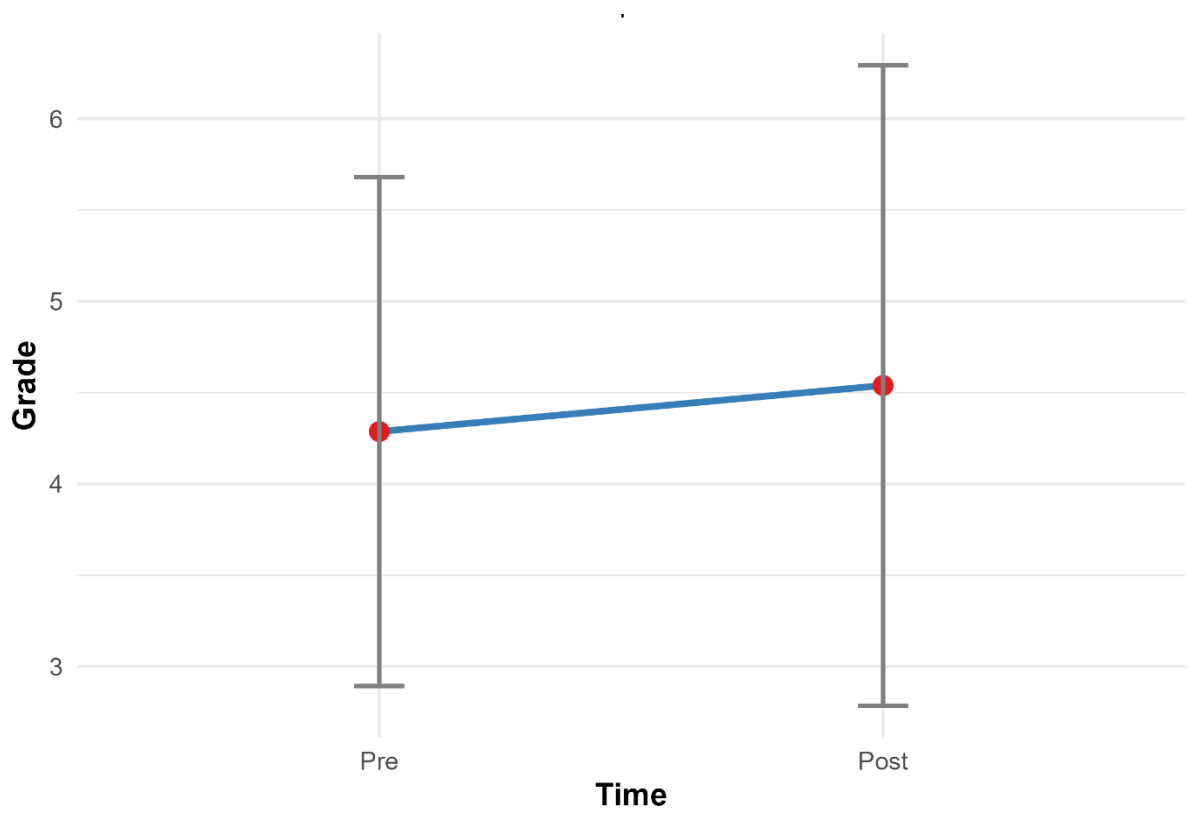

Figure 7: Mean grades before (predicted) and after (actual) the Arts Tutoring intervention.

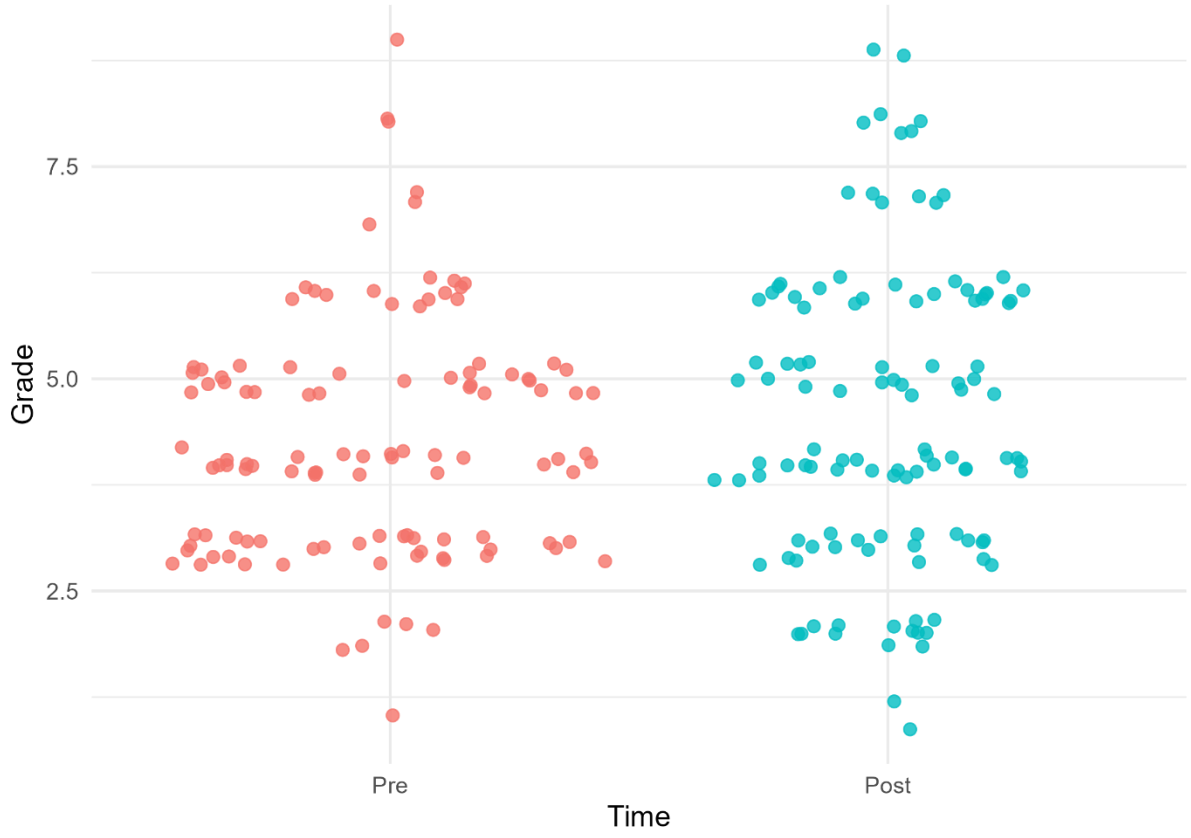

Figure 8: Distribution of individual grades before (predicted) and after (actual) the Arts Tutoring intervention.

Results indicated a small overall increase in grades from predicted to actual GCSE grades (see Figure 7). However, this improvement did not reach statistical significance ($p = 0.103$). While this suggests a positive trend, the limited sample size (n = 27) reduces statistical power, meaning that modest effects may not have been detected. As illustrated in Figure 8, the distribution of predicted and actual grades also shows a slight positive shift, supporting the overall pattern of improvement. Nevertheless, these findings cannot be confirmed statistically and should therefore be interpreted with caution. Increasing the sample size in future iterations could strengthen confidence in these trends (see Recommendations section).

A follow-up analysis was conducted to examine whether the number of intervention hours correlated with the improvement patterns observed in the previous analysis. Variation in intervention hours across students was due to several practical factors, including differences in school responsiveness and timeliness in sharing student information, student absences due to illness or other reasons and group revision due to lack of target students. To this aim, an LMM was performed, with time (pre-, predicted vs post-, actual grades), contact hours and their interaction as fixed effects and with individual differences between students' baseline performance and trajectories as random effects. By modelling these sources of variation, the LMM allows for an

estimation of the relationship between intervention dosage and observed improvements while handling the nested structure of repeated measurements within participants. This follow-up analysis revealed a final key finding:

KEY FINDING 3: There was a tentative, non-significant trend suggesting greater improvement with more intervention hours, though this should be interpreted cautiously given the small sample size.

The model reiterated the results obtained in the main Wilcoxon test above: it suggested a slight improvement in grades from predicted to actual, without reaching statistical significance ($p = 0.792$). The p-value from the LMM is considerably higher than that from the Wilcoxon test ($p = 0.103$), reinforcing the suggestion that the small observed increase could have occurred by chance rather than reflecting a real effect of the intervention. More interestingly, the analysis of the interaction between number of contact hours and grade changes showed a small, non-significant trend ($p = 0.287$), hinting that higher contact hours might be associated with slightly greater improvement. However, and to reiterate, the evidence is not conclusive and should be interpreted with caution, particularly given the small sample size.

Recommendations

1. **Increase sample size and/or boost student participation to ensure stronger data quality.** The current sample was too small to draw any meaningful conclusions from the analysis, which limits the usefulness of the evaluation. Moreover, although comparison group data was available from students who did not participate in the programme, the imbalance in group sizes (27 intervention vs 88 non-intervention) would make the inferential analysis unreliable.[2] To make the current methodology informative for this programme's evaluation, an increase in sample size would be required. If this is not feasible due to the nature of the programme, it would be advisable to adopt more suitable evaluation designs, as suggested in Recommendation 2 below.

2. **Consider small-n methodologies or qualitative case studies as the evaluation methodology, rather than the current before-and-after design.** Given the

[2] For this reason, the analysis focused primarily on within-group changes in the intervention group. In addition, the comparison group was not collected for evaluation purposes, hence the students were neither matched nor selected based on relevant characteristics, providing an additional reason for discarding this data.

programme's inherently small number of participants, increasing the sample size is likely impractical. Therefore, small-n evaluation methods, which are better suited to deal with smaller samples, will provide more meaningful insights and a more accurate evaluation of its impact. Additionally, incorporating teacher feedback could complement the small-n evaluation and grade comparisons, offering first-hand evidence of observable changes in classroom participation and performance and in line with recent guidance on monitoring and evaluating tutoring (DfE, 2024).

3. **Consider and incorporate an IAG component**. While findings show some positive impact on higher education aspirations, this remained the least impactful benefit. Given its one-to-one design, a more structured IAG strand could be embedded, for example, through exploring the links between the skills being developed and the subject (or similar ones) at A level, or through dedicated discussions of creative HE routes.

4. **Enhance the exploration of the impact of increasing intervention hours on grade improvement, particularly if sample size is increased.** The current evaluation did not find a statistically significant relationship between the number of intervention hours and grade improvement, though a small, non-significant trend suggested that students with more hours might experience slightly greater gains. This aligns with existing research indicating that frequent, shorter sessions (of up to an hour) tend to produce the greatest impact (EEF, 2022). Future evaluations should strengthen this analysis by examining the effect of the intervention dosage more systematically, ideally with larger sample sizes or more detailed individual-level data to assess whether additional hours have a measurable impact. Moreover, exploring whether the effect of contact hours differs across students with varying baseline performance or other characteristics could provide further insight into who benefits most from additional intervention.

References

Department for Education (2024). *Tutoring: guidance for education settings*. Tutoring: guidance for education settings

The Education Endowment Foundation (2022). *Making a difference with effective tutoring*. Making a Difference with Effective Tutoring | EEF