EFFICACY

Amplify Science field trial effectiveness research report

Amplify.

Table of contents

Executive summary

Amplify Science Middle School was field tested with more than 475 teachers and 34,000 students. For each unit, students responded to an assessment prior to instruction and again immediately following instruction. Each of these pre-unit and post-unit assessments were scored and student growth calculated using percent growth and effect size measures. Teacher feedback was collected via daily surveys. Here we present the results of the field trial analysis to understand how effective each unit was in achieving the learning objectives, and how teachers and students responded to the material. Given the results for each unit, the typical teacher might expect student knowledge to increase by 50% or more per unit, or roughly a 1- to 2-letter grade improvement. If given the choice, nearly all teachers surveyed would choose to use Amplify Science in their classrooms.

Amplify Science

Amplify Science is a brand new K–8 science curriculum, designed explicitly to meet 100% of the NGSS requirements. In Amplify Science, students are learning to think, read, write and argue like actual scientists by conducting investigations, creating and critiquing models and gathering evidence from a variety of sources.

This white paper addresses the effectiveness of the Amplify Science Middle School program as demonstrated by the developmental field trials that were conducted with teachers over a two-year period. The effectiveness of the program is determined through both the observed outcomes (i.e., student growth in learning) and the experience (i.e., teacher satisfaction).

Amplify Science is the result of a unique partnership between The Lawrence Hall of Science at UC Berkeley and Amplify Education, Inc.

Curriculum development process

Every unit undergoes a rigorous development and testing process before it becomes available for general implementation. This approach to field testing and improvement is a hallmark of curriculum development at the Lawrence Hall of Science (LHS). First, the education experts at LHS consult with external scientific experts in each topic's field of study. Based on expert input and a knowledge of the latest advances in the learning sciences, a first version of each unit is drafted. The Lawrence Hall of Science then conduct a series of small-scale development pilots (field trials) and gathers some initial impressions from teachers. Based on this feedback, the units are revised to create a second-round draft. This second draft of a unit could be considered complete, but at this point an even larger field trial is conducted in classrooms around the country. The large-scale field trials are meant to replicate the ways in which the program would be implemented at scale, in real classrooms. Each unit is taught start-to-finish in many different settings across the country. After a thorough analysis of the data and feedback gathered in the national field trials, LHS makes final revisions to the unit. Only after this last round of improvement and refinement is the curriculum ready for the classroom. The remainder of this paper discusses some of the findings from these field trials.

Our testing steps

1. We consult with external experts in each unit's field of study.
2. We conduct a developmental pilot.
3. We revise units based on feedback from piloting teachers.
4. We conduct field tests in classrooms around the country.
5. We make final revisions based on feedback from the field.
6. Our curriculum is ready for the classroom.

Field trial outcomes

The 2014–2016 field trial program

Starting in the 2014–2015 school year and concluding in 2015–2016, more than 475 teachers and 34,000 students in cities, suburbs, and rural communities across the country used Amplify Science in their classrooms as part of a developmental pilot. A total of 16 units spanning Earth, Life, and Physical sciences and nine units covering engineering and technology were conducted in partnership with middle schools.

Methods

Student growth and learning outcomes

For each science unit, a Pre-Unit Assessment and a Post-Unit Assessment was administered to students as an online survey using a pre-experimental, one-group design. This descriptive research design was chosen as the optimal approach for evaluating the developmental version of the units while balancing costs and resources despite the lack of external validity measures.

The two assessments were identical in content and therefore targeted a single scale for each unit. The assessments were comprised of 9 to 21 selected response items and two constructed response items. Teachers participating in the field trial program were instructed to administer the Pre-Unit Assessment on the very first day of class, before any instruction had occurred, and to give the Post-Unit Assessment immediately after the final lesson of the unit. The selected response items were scored dichotomously.

For measuring the effectiveness of each unit, we compared results from a subset of students for whom we had paired Pre-Unit and Post-Unit assessments. Effectiveness was measured in two ways: as percent growth between paired pre- and post-test scores and as practical effect size, known as Cohen's d (Cohen, 1988). Percent growth was calculated as the fraction of change in overall score compared to the initial score. Specifically,

$$growth = \frac{\bar{x} - \bar{x}_0}{\bar{x}_0} \times 100$$

where $\bar{x}$ is the average Post-Unit Assessment score and $\bar{x}_0$ is the average Pre-Unit Assessment score.

Cohen's *d* is also calculated using the effsize package (version 0.6.4) and the R statistical programming environment (version 3.3.1), using pooled standard deviations. Cohen's *d* is calculated as the difference between Pre-Unit Assessment mean score and Post-Unit Assessment mean score in terms of standard deviations, following the protocol outlined in Dunst et al., 2004.

Teacher surveys

For each science and engineering unit, we asked teachers to respond to a questionnaire before the unit, at the conclusion of each lesson, and at the very end of the unit. Questions were open-ended and rating-scaled. Likert-scaled questions were analyzed and results reported below as percentage of total respondents per response category.

Unit results

To gauge the magnitude of the growth in student knowledge, we looked at the percent growth between Pre- and Post-Unit assessments and the effect size of the growth using Cohen's *d*. The percent growth and effect size measures for all units are listed in Table 1. Many educators are familiar with Hattie's effect size scale (Hattie, 2008), but caution should be used when comparing these results to that scale, as there was no control group comparison in this single-case design. Our average effect size of 0.84 and the average percent growth was 45% across all units. Many units had an effect size that suggests students grew in knowledge more than a single standard deviation, indicating highly practical and successful learning outcomes. This is confirmed by the observations that most students moved up several letter grades after having completed the unit.

Table 1. Growth results

Unit	Number of students	Percent growth	Effect size (Cohen's *d*)
Chemical Reactions	722	66.80	1.28
Earth's Changing Climate	486	42.92	0.70
Plate Motion	786	42.22	0.87
Evolutionary History	483	27.81	0.47
Force and Motion	838	40.09	0.67
Light Waves	663	53.53	1.01
Magnetic Fields	680	67.61	1.30
Matter and Energy in Ecosystems	337	27.51	0.76
Natural Selection	704	13.82	0.47
Ocean, Atmosphere, and Climate	555	32.87	0.74
Phase Change	207	105.96	1.33
Populations and Resources	95	6.11	0.13
Thermal Energy	1299	35.77	0.82
Traits and Reproduction	973	54.55	0.85
Weather Patterns	761	63.92	1.18
Field trial average	**639**	**45.43**	**0.84**

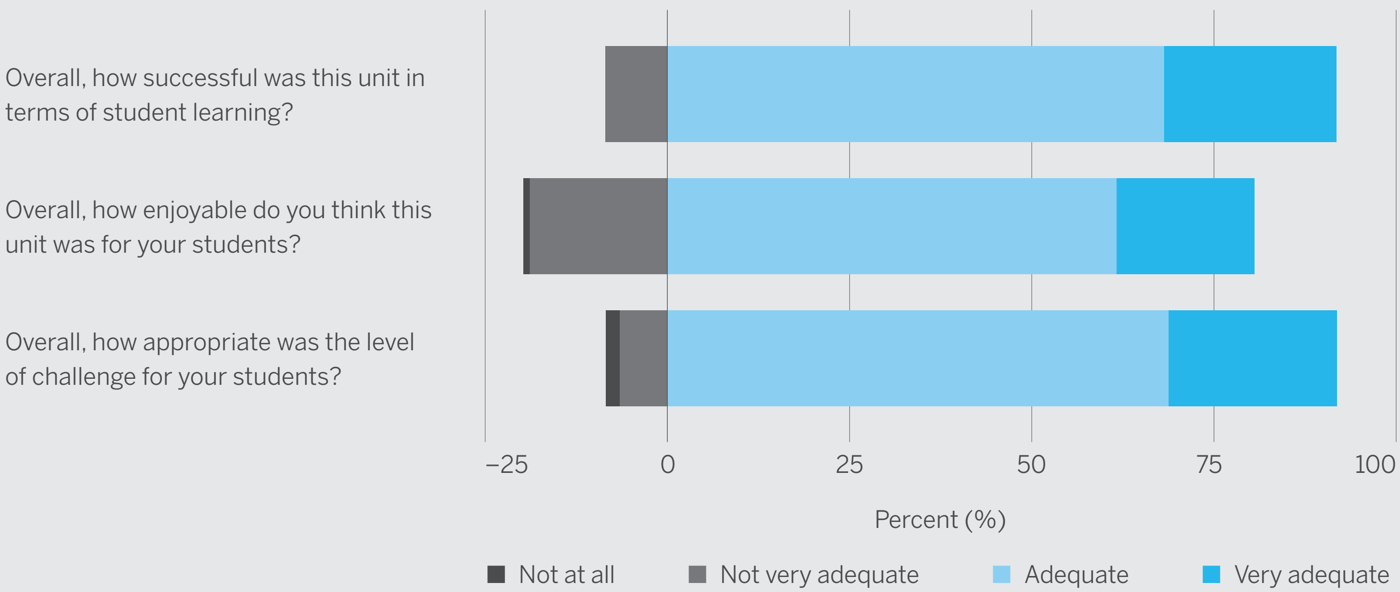
Teacher survey results

In addition to measuring student learning, we also gathered feedback from teachers throughout the implementation of each unit. The results of the final survey given to teachers at the conclusion of each field trial are presented below.

Teacher & student satisfaction

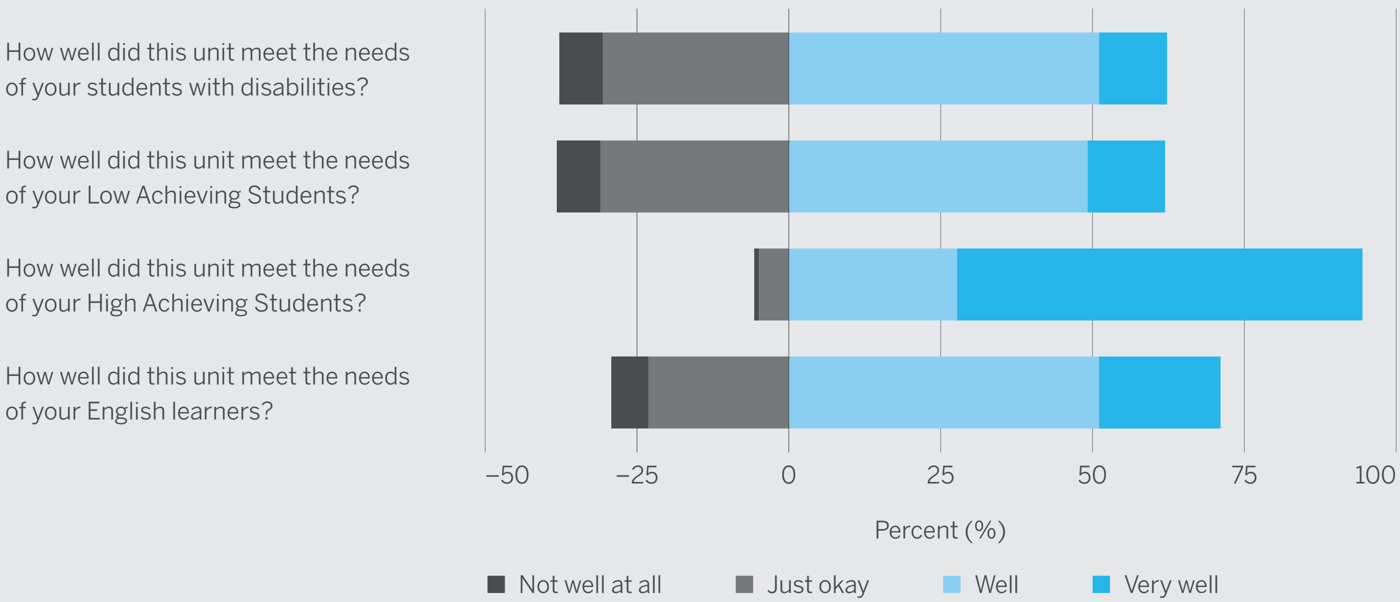
After concluding the field trials, teachers were satisfied with the units enough to use the curriculum again. More than 85 percent of our field trial teachers would teach with Amplify Science again. Students were also satisfied with the curriculum, as reported by their teachers. Figure 2 (Student impressions & outcomes) shows the breakdown of student impressions, with more than 80 percent of teachers reporting that the lessons were successful, enjoyable and appropriately challenging for their students.

Figure 2. Student impressions & outcomes



Breaking down the student impressions by different populations of learners in Figure 3 (**Meeting student needs**), we see that teachers thought that Amplify Science met the needs of their advanced learners, lower level learners, students with disabilities, and English learners.

Figure 3. Meeting student needs



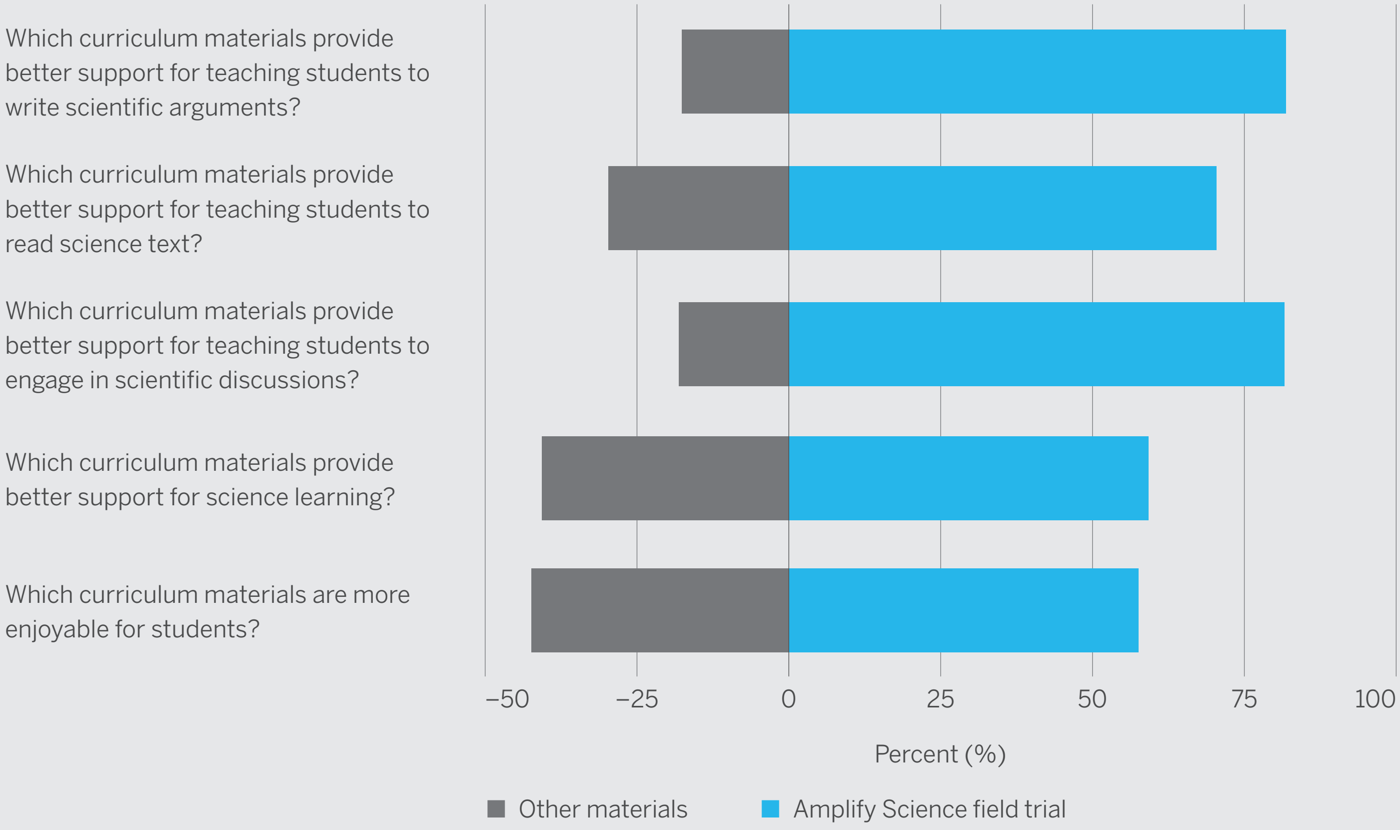
Effectiveness of curriculum materials

When asked to rate the unique features of Amplify Science and compare it to their usual curriculum materials, teachers indicated a strong preference for Amplify Science. Compared to the materials usually used in the classroom, teachers said the following about Amplify Science:

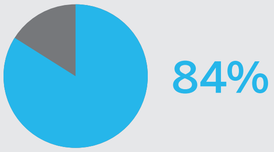
- **2 : 1** teachers say Amplify Science **supports science learning better**
- **2 : 1** teachers say Amplify Science is **more enjoyable for students**
- **2 : 1** teachers say that Amplify Science **provides better reading support**
- **5 : 1** teachers say the Amplify Science is **better at engaging students in scientific discussions**
- **5 : 1** teachers say that Amplify Science is **better at supporting students to write scientific arguments.**

Amplify Science is unique in the way each unit connects students' learning to a real-world phenomenon, and 84 percent of teachers reported that their students reacted well or very well to these phenomena. The investigative nature of each unit is also supported by the interactive simulations and digital modeling tools that are used in most units.

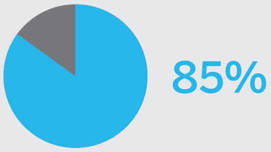
Figure 4. Comparison with other curricular material



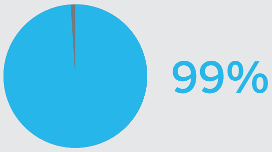
3 out of 4 teachers said Amplify Science did a good to very good job of meeting their student's diverse needs.



84% of students reacted well or very well to the real-world phenomana.



85% of teachers would use the Unit again.



99% of teachers thought the simulations were helpful, enjoyed by students, and appropriately complex.

Figure 5. Curriculum features: Modeling Tools

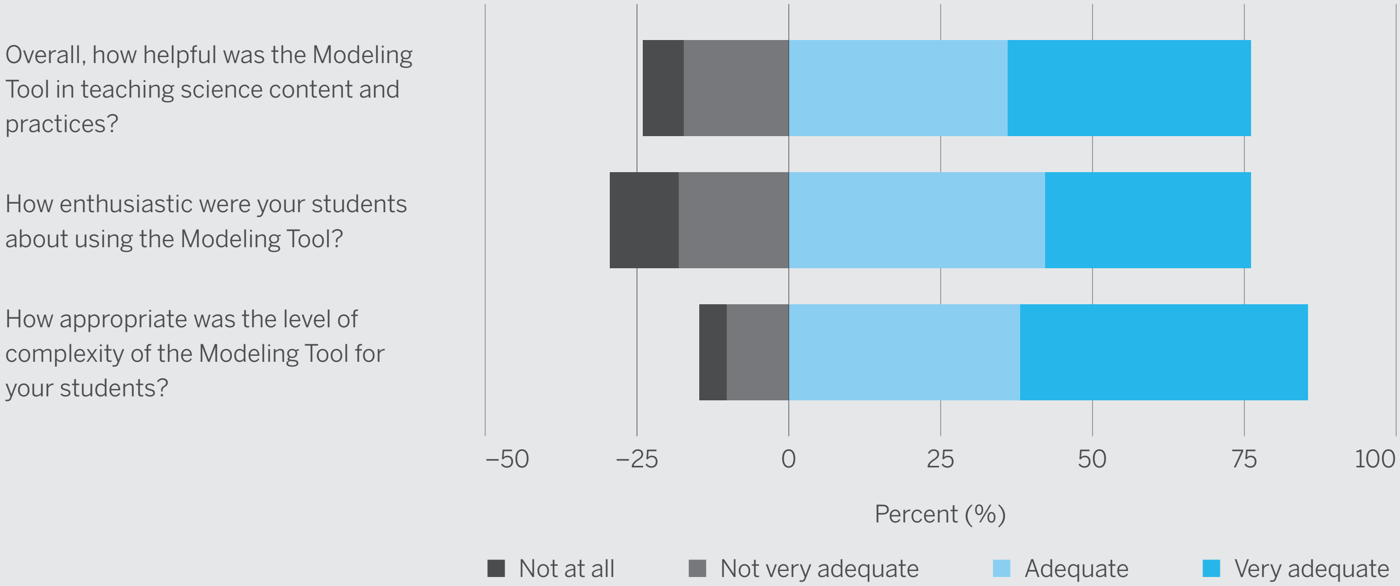
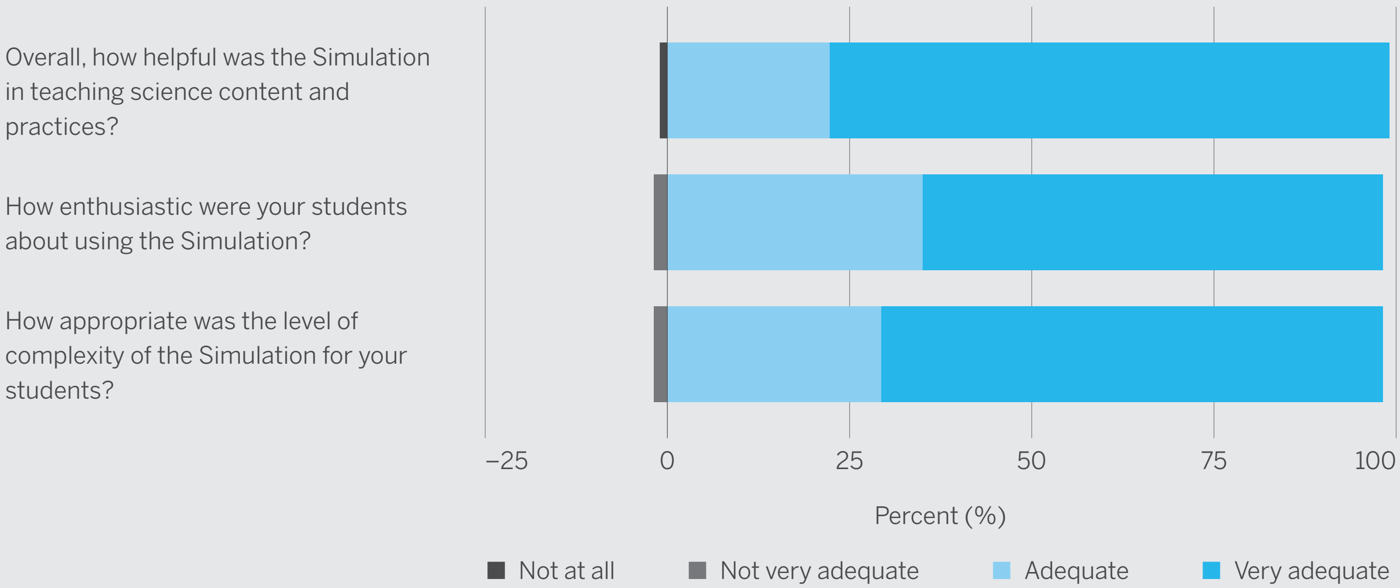


Figure 6. Curriculum features: Simulation



Summary

Following guidelines for educational contexts, the effect sizes observed were always above average (with the single exception of the Populations and Resources unit). Typical effect sizes ranged from 0.4 to over 1, indicative of a highly successful program. Given the observed Cohen d values and the percent growth for each unit, the typical teacher might expect student learning to increase by 50 percent or more per unit, or roughly a one- to two-letter grade improvement.

Teacher feedback consistently indicated that a large majority of teachers would use the program again if given the choice, thought the program was effective and enjoyed by students, addressed the needs of diverse learners, and was generally better than other curricular material they have used in the past. In particular, the simulations and the real-world phenomena used in the units were reported to be well received by teachers and students.

How we used these results to improve the program

These measures of student growth were based on teachers' and students' first use of each unit as the material was being developed. Units have been modified, extended, and generally improved since the time that these field trials were conducted. In addition to the quantitative measure of student growth, we also considered teacher feedback collected via daily surveys, measures of student progress against unit learning goals, and the psychometric properties of the assessment items themselves. All this information was reviewed as part of the curriculum development process and led directly to the current version of Amplify Science.

References

Cohen, J. (1988). Statistical power analysis for the behavioral sciences. (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Dunlap, W. P., Cortina, J. M., Vaslow, J. B., & Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. Psychological Methods, 1, 170–177.

Dunst, Carl J.; Hamby, Deborah W.; Trivette, Carol M. (2004) Guidelines for Calculating Effect Sizes for Practice-Based Research Syntheses. Centerscope. Vol. 3:1, 1–10.

Hattie, John. (2008). Visible Learning: A Synthesis of over 800 Meta-Analyses Relating to Achievement. London: Routledge.

For more information visit amplify.com

Corporate:
55 Washington Street
Suite 900
Brooklyn, NY 11201-1071
(212) 796-2200

Sales Inquiries:
(866) 212-8688 • amplify.com

Amplify.