MathVL-test Accuracy Across Dataset Sizes