Research Brief
Categorical Growth Analysis: A Method Involving the Repeated-Measures T-Test

by Sarah Callahan, Ph.D.
Assessment Technology Incorporated

Overview: As part of the growing implementation of instructional effectiveness initiatives, districts and schools have expressed a need for information about the growth of the students for which each teacher and principal is responsible. To be useful, student growth must be estimated based on valid, reliable student assessments and categorized via robust statistical approaches. Via the Galileo® K-12 Online Instructional Improvement and Instructional Effectiveness System (IIIES), Assessment Technology Incorporated (ATI) provides districts/schools with a wide variety of valid, reliable assessments including instructional effectiveness (IE) pretests and posttests as well as integrated statistical approaches designed to provide districts/schools with precise information about student growth.

This research brief describes and illustrates the underlying method for one of the statistical approaches developed by ATI to provide information about student growth, Categorical Growth Analysis. Categorical Growth Analysis enables educators to evaluate growth throughout the year. Information about student growth can be accessed along with information about student achievement in Galileo reports and used to guide professional development, student intervention, and other activities. Categorical Growth Analysis supports a fair, constructive approach to educator evaluation that evaluates educators against a defensible standard while providing every educator with the opportunity to succeed. This brief also summarizes the results of a simulation study evaluating the appropriateness of the method underlying Categorical Growth Analysis for a wide range of possible datasets.

Method: Categorical Growth Analysis considers the relationship between the growth observed for a group of students over a given time period and the growth expected for those students in that time period. Categorical Growth Analysis employs a well-established common statistical test, the repeated-measures t-test, to evaluate whether the difference between observed growth and expected growth is significant for the group of students for which a teacher/principal is responsible.

Categorization of Student Growth: A classification of Expected Growth Exceeded indicates that observed growth was significantly higher than expected growth while a classification of Expected Growth Not Maintained indicates that observed growth was significantly lower than expected growth. A classification of Expected Growth Maintained indicates that observed growth was not significantly different from expected growth.

Calculation of Observed Growth: Each student's ability is measured at two points in time (yielding two Developmental Level [DL] scores) and the difference between the DL scores represents the observed growth. DL scores are generated via procedures based in Item Response Theory and placed on a common scale to support the evaluation of growth. For educator evaluation at the end of the year, DL scores on an instructional effectiveness pretest and an instructional effectiveness posttest are compared. Throughout the year, DL scores on an initial district-wide assessment and a subsequent district-wide assessment can be compared to generate an interim categorization.

Calculation of Expected Growth: ATI provides districts/schools with annually updated estimates of the expected growth for students in a wide variety of grades and content areas. These estimates are based on research conducted across all ATI clients and currently represent the expected growth used as part of Categorical Growth Analysis. Future development will also enable districts/schools to establish their own standards for expected growth if desired.

Theoretical Illustration: The repeated-measures t-test evaluates whether the difference between mean observed growth and expected growth is so large it is unlikely to have occurred by random chance. Figure 1 presents a theoretical illustration of the results of the repeated-measures t-test in three scenarios corresponding to the three possible categorizations in Categorical Growth Analysis. Within each graph, the black curve represents a theoretical underlying distribution of observed growth scores for a group of students. The red line represents mean observed growth for the group and the blue line represents expected growth. The light red box represents possible mean observed growth values that might have been observed by random chance given the underlying distribution.

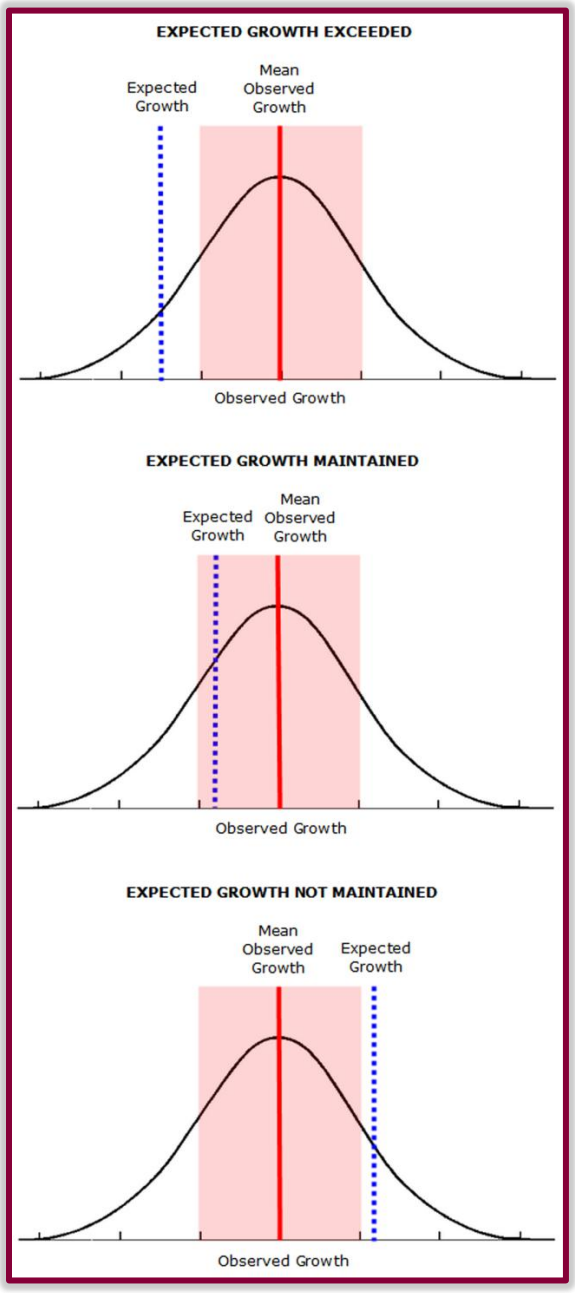

Figure 1
Theoretical Illustration

In the top graph, the blue line is outside the light red box and the t-test reveals that the difference between mean observed growth for this group of students and expected growth is unlikely to have occurred by chance. In other words, the difference is significant. In this case, the blue line is to the left of the red line meaning observed growth is higher than expected growth and the group of students is classified as Expected Growth Exceeded. In the bottom graph, the blue line is also outside the light red box, indicating a significant difference. However, the blue line is to the right of the red line meaning that observed growth is lower than expected growth and the group of students is classified as Expected Growth Not Maintained. In the middle graph, the blue line is inside the light red box. This means observed growth is not significantly different from expected growth and the group of students is classified as Expected Growth Maintained.

Sample Demonstration: This section presents a sample demonstration of Categorical Growth Analysis. The sample dataset in Table 1 illustrates DL scores on the IE pretest and IE posttest for the ten students in a hypothetical fifth grade math class taught by Mrs. Smith. Observed growth for each student was calculated by subtracting the student's DL score on the IE pretest from their DL score on the IE posttest.

TABLE 1
Sample Dataset

Ms. Smith's 5th Grade Math Class			
Student	**DL Score 1 (IE Pretest)**	**DL Score 2 (IE Posttest)**	**Observed Growth (DL2-DL1)**
1	975	1055	80
2	998	1074	76
3	1011	1110	99
4	984	1071	87
5	1009	1095	86
6	1053	1170	117
7	972	1097	125
8	955	1059	104
9	1024	1109	85
10	968	1034	66

In the sample dataset, the mean of the observed growth values is 92.5. In this sample, the IE pretest was administered September 1 and the IE posttest was administered April 1. Based on ATI's research, the expected growth for a student in fifth grade math across this time period is 78. The repeated-measures t-test evaluates whether the mean observed growth (92.5) is significantly different from the expected growth (78) considering the variability in the observed growth values and the number of students in the class. In this sample, the t-test revealed that observed growth was significantly higher than expected growth, yielding a classification of Expected Growth Exceeded.

Simulation Results: ATI conducted a simulation study to evaluate the robustness of the repeated-measures t-test as a method for Categorical Growth Analysis. The study involved 84,672 simulated datasets. Each simulated dataset included DL scores at two time points for a group of students. As part of the simulation, a number of critical parameters were systematically varied in order to generate datasets that would represent the range of possible scenarios likely to be encountered in real datasets.

These critical parameters included the size of the sample of students (from 5 to 100 students) and the magnitude of the difference between observed and expected growth (from 50 pts lower than expected to 80 pts higher than expected). The simulation study revealed that the repeated-measures t-test was extremely robust across a wide range of scenarios involving samples of 10 or more students. The study also revealed that the repeated-measures t-test was sufficiently sensitive for the desired purpose as significant differences between observed and expected growth were detected in a wide range of scenarios.

Conclusion: The repeated-measures t-test method for Categorical Growth Analysis provides a defensible statistical approach to evaluating the growth of the students for which a teacher and/or principal is responsible. ATI's research demonstrates that this approach is robust in a wide range of scenarios. Categorical Growth Analysis is also associated with a number of practical benefits. First, it can be used to evaluate growth on an ongoing basis throughout the year, enabling educators to use this information to guide instruction and intervention. Second, it considers each educator separately in comparison to a standard. Since educators are not compared to each other, it is possible for all educators to achieve positive results. Third, it focuses on student growth rather than student achievement, providing a source of complementary information about the effectiveness of instruction. For example, Categorical Growth Analysis can identify groups of low performing students who have displayed exceptional growth as well as groups of high performing students who have displayed inadequate growth. Given the observed properties as well as the practical benefits, the repeated-measures t-test method for Categorical Growth Analysis is a valuable tool that will contribute to a district/school's efforts to improve instructional effectiveness and elevate student achievement.