



Analysis of Test Characteristics



Prepared by:

Michelle Watts, Ph.D.

Psychometrician

March, 2016

Table of Contents

Introduction	3
Table 1: Student Results.....	3
Item Analysis	4
Figure 1: Distribution of Grade 3 item P values	4
Figure 2: Distribution of Grade 4 item P values	5
Figure 3: Distribution of Grade 5 item P values	5

Introduction

Progress Testing created formative assessments that were parallel to the Florida Standards Assessments (FSA) for grades 3 through 5 ELA and Mathematics using items exclusively from our Florida Formative Assessment Bank (FL FAB). The tests were administered to approximately 350 students attending low-achieving schools in Broward County, Florida. This report was created to analyze the psychometric characteristics of the tests, including the student achievement results and reliability estimates.

The student results are presented in Table 1. The mean percent correct for the ELA tests was 46.4%, 45.5%, and 50.2% for grades 3, 4, and 5, respectively. The tests were difficult for the students, but it should be noted that the students who were administered the formative tests attended lower-achieving schools.

The mathematics tests were also very difficult. The mean scores, as shown in Table 1, were 51% correct for grade 3, 48% for grade 4, and 43% for grade 5. The grade 5 test included a variety of technology-enhanced items. Some of those items required the student to select multiple correct answers. If the student selected one of the correct answers but not others, the item was marked incorrect. There was not credit given for partially correct answers. There were also numeric response items in which the student was required to produce the answer to the question. These items did not allow for guessing, which would also reduce the students' percent of correct answers. These item types are possible explanations for the reduced score for grade 5.

Table 1: Student Results

	N	Mean	Standard Deviation	Cronbach's Alpha
Grade 3 ELA	337	46.4%	17.2	.9131
Grade 4 ELA	426	45.5%	16.1	.9082
Grade 5 ELA	344	50.2%	16.9	.9051
Grade 3 Mathematics	327	51.1%	16.6	.8849
Grade 4 Mathematics	313	47.6%	17.9	.8996
Grade 5 Mathematics	427	43.1%	18.7	.9161

Test reliability is a very important characteristic used to judge the quality of any test. Reliability is the consistency which students responded to the test items. Cronbach's alpha was used to estimate the reliability for these formative assessments. As a measure of internal consistency or the inter-relatedness of the items, reliability estimates should be greater than .70.

As shown in Table 1, the reliability estimate for the grade 3 ELA test was .91, indicating a very high level of reliability. Grades 4 and 5 ELA tests were also found to have very high estimate of reliability at approximately .91 for each test.

The reliability estimates for the mathematics formatives were slightly lower than the ELA tests at .88 for grade 3, .90 for grade 4. The grade 5 mathematics test had the highest reliability at .92. The reliability of the ELA and mathematics formative tests can be considered extremely high.

These formative tests were created to provide educators with feedback on student progress toward meeting the standards measured on the state-administered FSA. Predictive validity on the tests will be established when student FSA scores become available.

A review of the P value (percent correct) for each item will provide information on the overall difficulty of the test with more detail than the mean score alone. Figure 1 presents the distribution of P values for the ELA and mathematics Grade 3 test. Both tests can be described as difficult for the students. The ELA test contained very few items that the students found easy (P values $\geq .80$). The majority of the items can be classified as very difficult with item P values $\leq .60$. The P values alone do not indicate high- or low-quality test items.

Item Analysis

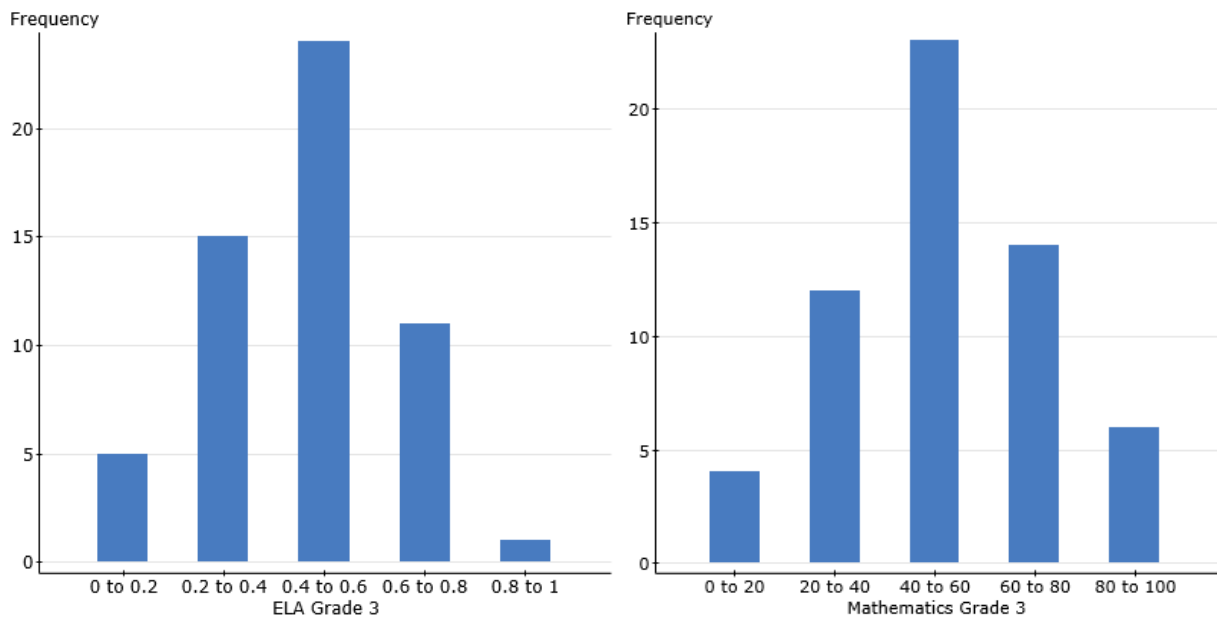


Figure 1: Distribution of Grade 3 item P values

Figure 2 displays the P values for the ELA and mathematics tests for grade 4. These tests were also found to be difficult containing many items with P values below .6. While the grade 5 test contained a greater number of easy items, the addition of more easy items to both tests is suggested. The same suggestion is made for the grade 5 tests.

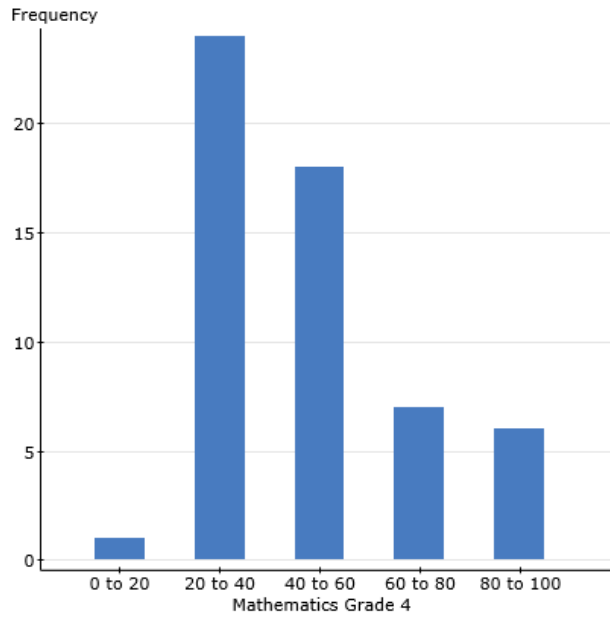
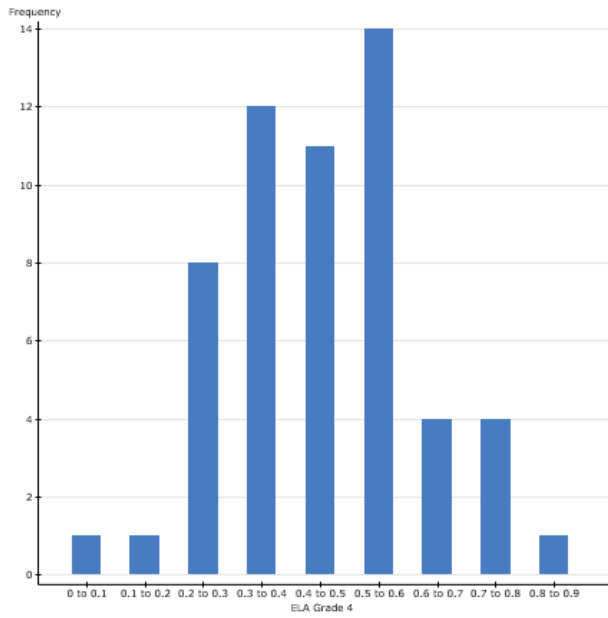


Figure 2: Distribution of Grade 4 item P values

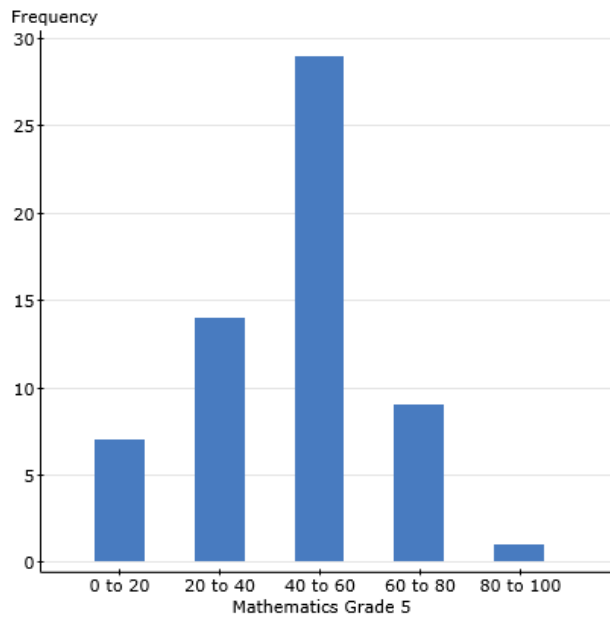
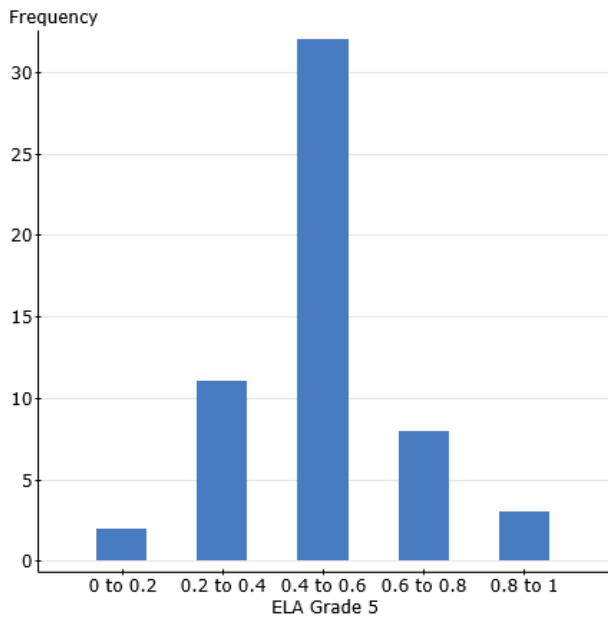


Figure 3: Distribution of Grade 5 item P values