Evaluation and measurement play an important role in helping educators cope with their environment of teaching and learning programs. Evaluation involves judging the value of or worth of a pupil, an instructional method, or an educational program. To make valid judgments, an evaluator needs to conduct measurement through testing. As an important measuring instrument, the test of UN should have a high quality and be able to give reliable and accurate measurement results. The quality of a test can be identified by knowing its difficulty level and the discrimination index of the test.

The purpose of the study was to investigate the difficulty level and the discrimination index of English test items in UN 2005/2006 which was readministered to the third grade students of MA Al-Huda Wajak, Malang. The descriptive design was used as no treatment was given to the sample. In this study, the population was all the third-grade students (110) of MA Al-Huda. The sample was taken by using cluster random sampling that was natural science class. The procedure of collecting the data was using the English test items in UN 2005/2006. The research was done by using some steps as follows: 1) administering the test using test of UN 2005/2006; 2) collecting all students answer sheets; 3) scoring all students answer; 4) arranging the answer sheets to sequence high to low score; 5) dividing the students into two equal groups, upper and lower scoring group; 6) analyzing the test items.

The result of analyzing the data was in the form of descriptive statistics by using the formula of difficulty level and discrimination index. The result of this study showed that the average value of difficulty level was 0.43 (moderate) and the average value of discrimination index were 0.23 (poor). It means that test items did not fulfill Azwar’s criteria of a good test that a test should have difficulty level at around 0.50 and discrimination index at 0.30 up to 1.00. In other words, the English test item of UN 2005/2006 was less qualified.