Many investigators have reported attempts to develop reliable laboratory and clinic evaluation systems. However, few studies, regardless of level of success, have used an analytic procedure to identify those components of the evaluation system that, if refined further, could improve reliability. The purpose of this study was to compare intra- and interexaminer variability in two evaluation methods: glance and grade (global), and checklist and criteria (analytical). Three faculty staff members with more than ten years of clinical and teaching experience evaluated operative procedures performed on plastic teeth representing the primary teeth by thirty dental students in Pediatric Dentistry preclinical laboratory sessions. The preparations were graded blindly by each of the three evaluators (A, B, and C) three times without magnification. The values were statistically analyzed using Wilcoxon signed ranks test and Friedman test setting value of significance at 5%. The study revealed that among the three examiners, the intraexaminer variability was non significant in most situations. On the other hand, there was statistically significant variability between evaluators (i.e., interexaminer) for almost all preparations. Neither cutting off the scores nor using either evaluation methods (glance and grade or criteria and check list) caused an improvement in variability. The problem of interexaminer reliability and variability still existed.