Reliability of the Advanced Dundee Endoscopic Psychomotor Tester for Bimanual Tasks

Nader K. Francis, MB, ChB; George B. Hanna, PhD; Alfred Cuschieri, MD, FRSE

**Objective:** To evaluate the reliability of the Advanced Dundee Endoscopic Psychomotor Tester (ADEPT).

**Design:** The Advanced Dundee Endoscopic Psychomotor Tester was developed for objective evaluation of bimanual endoscopic tasks. The system is in several aspects relevant to an actual endoscopic environment, and initial studies showed a strong correlation with clinical competence. Twenty medical students were tested on ADEPT (10 runs in 2 sessions). Their performances for 2 sessions were analyzed by the Spearman $\rho$ correlation to examine test-retest reliability. Coefficient $\alpha$ was used to indicate the internal consistency of the system.

**Results:** There was no significant improvement in task performance during the 10 runs. A positive correlation on ADEPT performance was found between the 2 test sessions. A coefficient $\alpha$ of .98 was observed between the different tasks of ADEPT.

**Conclusion:** These findings confirm that ADEPT is a reliable system for assessment of bimanual endoscopic task performance.

Arch Surg. 2001;136:40-43

Computer-based assessment methods are being developed as means of objective assessment of task performance. They provide standard tasks for valid comparisons between subjects and at different times. The face validity of these systems depends on the similarity of the test to the endoscopic environment and instrumentation. The design and configuration of these systems should permit objective assessment of performance in real time with minimal costs.

The Advanced Dundee Endoscopic Psychomotor Tester (ADEPT) was developed for objective assessment of bimanual task performance. The system provides a standard method of test application since all subjects are assessed in terms of the nature of the test and the instructions displayed on the computer screen. Aspects of face validity of ADEPT include (1) using a real endoscopic imaging system, (2) use of standard endoscopic instruments, and (3) movement of instruments inside a gimbal mechanism, allowing the same degrees of freedom as endoscopic instruments through access ports. The system was shown to be clinically valid by the correlation of ADEPT performance scores with independent assessment of clinical competence.
**SUBJECTS, MATERIALS, AND METHODS**

Twenty medical students participated in the study. Each subject performed 2 practice runs with ADEPT and then 10 test runs in 2 sessions (5 runs in each session). Each run entailed 20 tasks (each of the 5 tasks was performed 4 times). The sequence of tasks was randomly assigned by the system software. The medical students were from all years of medical school at the University of Dundee and had no previous surgical experience or orientation on ADEPT prior to the experiment. The subjects had a corrected visual acuity of at least 20/30 when tested using a Snellen chart.

**THE ADVANCED DUNDEE ENDOSCOPIC PSYCHOMOTOR TESTER**

The Advanced Dundee Endoscopic Psychomotor Tester consists of a dual gimbal mechanism that accepts standard endoscopic instruments to manipulate various tasks in the target object (Figure 1). The target object consists of a sprung base plate with 5 positioning tasks and a sprung top plate with access holes. The tasks consist of 2 sliders, a joystick, a dial, and a toggle switch. Each task entails manipulation of the top plate with 1 instrument to enable another instrument to negotiate the task in the back plate through the access hole. Instructions for the subject are displayed on the computer screen. The system registers an instrument error when the instrument comes in contact with the sides of the front plate holes.

The system employs a 2-dimensional video endoscopic system (Karl Storz, Tübingen, Germany) which consists of a Hopkins II 30° endoscope (model 26033BP) and Telecam single chip camera (model 20210030). Light within the dome of ADEPT is provided by a halogen light source (coldlight fountain 450V) connected to the endoscope with a fiberoptic light cable (model 495NB) and the image is displayed on a Sony monitor (model PVM-14043MD; Sony, Tokyo, Japan). Two alligator forceps are mounted in the gimbal mechanism. For the purpose of the study, the system was adjusted to provide a 60° manipulation angle (between the 2 instruments), a 30° azimuth angle (between the instrument and optical axis of the endoscope), and a 60° elevation angle (between the instrument and horizontal plane). The optical axis-to-target view angle was 85°. These angles provide the best ergonomic set-up for optimum task performance.5 The system software allowed an allocation time of 10 seconds to complete each task and an accuracy tolerance limit of 0.1 seconds for task positioning.

The end points of the system were the execution time to complete the task, instrument error contact time, flight trajectory of each probe (ie, x, y, z, and φ coordinates), and the number of successful tasks within the allocated time and tolerance limit.

**STATISTICAL ANALYSIS**

The data relating to execution time, instrument contact error time, angular deviations, and the success score of ADEPT were analyzed using the nonparametric Kruskal-Wallis test. Statistical significance was set at 0.05%. The nonparametric Spearman ρ correlation coefficient enabled the computation of the correlation between 2 sessions. This was undertaken to examine test-retest reliability. Coefficient α was used to test the internal consistency of ADEPT. It was obtained from the mean reliability coefficient from all possible split halves of the test outcome measures. Coefficient of variation was used to study the discriminative ability among subjects on ADEPT. It was derived by dividing SD over the mean and multiplying by 100.

**RESULTS**

The Table presents the median and interquartile ranges of the outcome measures of ADEPT for the 2 test sessions.

**TEST-RETEST RELIABILITY OF ADEPT**

In comparing the performance between the 2 sessions, no statistically significant difference was found in the instrument error, execution time, or angular deviations between the sessions and between the different runs in each session (P>.1). The success score reached the plateau after the second run in the first session, with no significant improvement thereafter.

The performance on ADEPT was correlated between the 2 sessions. There was a positive correlation of 0.64 for the execution time, 0.61 for the instrument error, and 0.60 for the success score (Figures 2, 3, and 4).

**INTERNAL CONSISTENCY OF ADEPT**

Internal consistency of the system was studied, using the coefficient α formula. This formula involves computing the correlation among all items (the 5 different tasks) of ADEPT and computing the average of the correlation. Tasks of the third run were analyzed to study the internal consistency among them. This showed a high coefficient α of .971 intertasks (between the 5 different tasks) and .979 intratasks (within the repeat number of each task) of ADEPT.

**DISCRIMINATIVE ABILITY OF THE ADEPT**

The discriminative ability among subjects on ADEPT was studied using the coefficient of variation, which is the SD-based measure of variability. The instrument error of ADEPT showed a high degree of variability, with a coefficient of variation of 68%. The success score of ADEPT had a coefficient of variation of 21.64%, while execution time showed only 13.76%. There was also a significant difference between subjects in all the outcome measures of ADEPT with P<.001.

**COMMENT**

The study confirms the reliability of ADEPT as an assessment method of endoscopic task performance. Several factors have been considered in studying this...
reliability. The participants undertaking the test were medical students who had no previous experience in endoscopic surgery. Hence, they were suitable to study the effect of practice on task performance and to examine test-retest reproducibility of the system. While a significant difference was found among subjects, which eliminates the possibility of tasks being either too easy or too difficult, the study showed no significant improvement with practice. This indicates the ability of ADEPT to demonstrate the difference among subjects at certain stages of their training and at the same time to allow for the inconsistency of human performance over time. In addition, the contrast validity of the system has been demonstrated since the system was able to distinguish between senior and junior surgeons and therefore can assess trainees’ progress with practice.6

Another factor that affects a reliability study is the method used to estimate reliability. The test-retest method assesses the degree to which test scores are consistent from one administration to the next by the same group of participants. The correlation between the 2 sets of scores of the 2 sessions on ADEPT was 0.6 for both error rate and execution time, and the success score. It is often impossible to consider the second administration of a test as an absolute parallel measure to the first in any psychological test because of the inconsistency of human performance and the carryover effect, despite the 2 sessions being carried out under standard circumstances. For these intrinsic problems in test-retest methodology, coefficient α, as a measure of internal consistency, was also used in the study. This method represents the most

<table>
<thead>
<tr>
<th>Outcome Measures</th>
<th>Session 1</th>
<th>Session 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Execution time, s</td>
<td>161.1 (22.9)</td>
<td>157.1 (25.9)</td>
</tr>
<tr>
<td>Instrument error contact time, s</td>
<td>15.4 (13.1)</td>
<td>15.5 (9.8)</td>
</tr>
<tr>
<td>Success score</td>
<td>13 (5.0)</td>
<td>14 (3.0)</td>
</tr>
<tr>
<td>Left x deviation (unit)</td>
<td>184.5 (286.7)</td>
<td>146.6 (269.6)</td>
</tr>
<tr>
<td>Left y deviation (unit)</td>
<td>226.7 (426)</td>
<td>169.6 (312.7)</td>
</tr>
<tr>
<td>Left z deviation (unit)</td>
<td>106.9 (166.9)</td>
<td>108.4 (194.3)</td>
</tr>
<tr>
<td>Right x deviation (unit)</td>
<td>340.2 (253.1)</td>
<td>272.5 (209)</td>
</tr>
<tr>
<td>Right y deviation (unit)</td>
<td>369.4 (251.2)</td>
<td>322.8 (183.7)</td>
</tr>
<tr>
<td>Right z deviation (unit)</td>
<td>198.6 (132.4)</td>
<td>174 (108)</td>
</tr>
</tbody>
</table>

*Data are given as median (interquartile range).
widely used and most general form of internal consistency estimate. A coefficient \( \alpha \) of .971 was found between the different tasks of ADEPT and a coefficient \( \alpha \) of .979 was found between the repeat of each task. This indicates very high internal consistency of the system in terms of the correlation among the different tasks and the reliability over the number of repetitions of each task in a run.

We are grateful to the medical students who participated in the study.

Corresponding author and reprints: Sir Alfred Cuschieri, Department of Surgery, Ninewells Hospital and Medical School, Dundee DD1 9SY, Scotland (e-mail: a.cuschieri@dundee.ac.uk).

REFERENCES